UNIT I FUNDAMENTALS OF ANALOG COMMUNICATION

1.2 Principles of Amplitude Modulation

The modulating signal modulates amplitude, frequency or phase of the carrier according to its variations in amplitude. This results in amplitude, frequency or phase modulation. The frequency and phase modulation is also called angle modulation.

1.2.1 AM Envelope and Equation of AM Wave

In amplitude modulation, the amplitude of a carrier signal is varied according to variations in the amplitude of modulating signal. Fig. 1.2.1 shows the modulating signal in Fig. 1.2.1 (a) Fig. 1.2.1 (b) shows high frequency carrier and Fig. 1.2.1 (c) shows amplitude modulated signal.

In Fig. 1.2.1 (c), observe that the carrier frequency remains same, but its amplitude varies according to amplitude variations of the modulating signal.

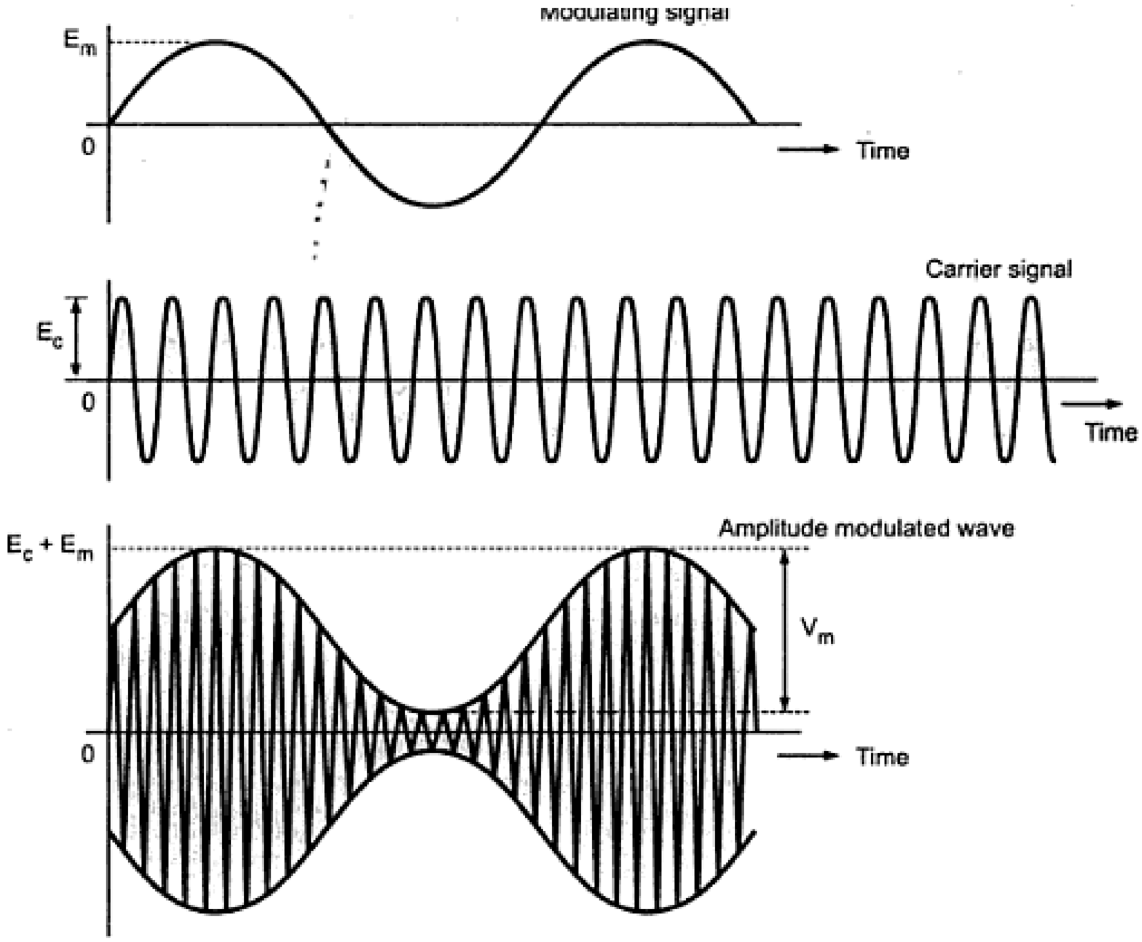


Fig. 1.2.1 (a) Sinusoidal modulating signal
(b) Sinusoidal high frequency carrier
(c) Amplitude modulated signal

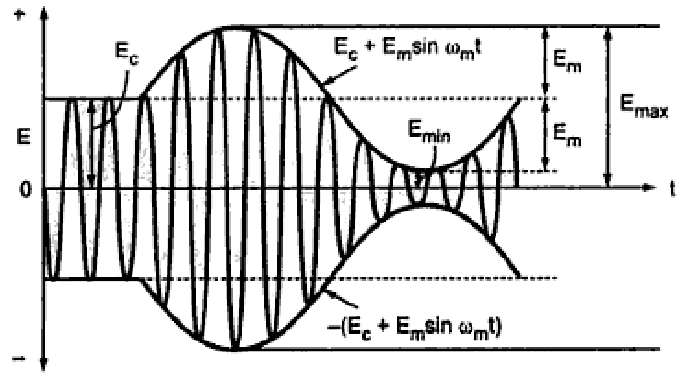


Fig. 1.2.2 AM wave

It is clear from the above signal that the modulating signal rides upon the carrier signal. From above figure we can write,

$$E_m = \frac{E_{max} - E_{min}}{2} \qquad \text{... (1.2.5)}$$

and

$$E_c = E_{max} - E_m \qquad \text{... (1.2.6)}$$

$$= E_{max} - \frac{E_{max} - E_{min}}{2} \text{ by putting for } E_m \text{ from equation (1.2.5)}$$

$$= \frac{E_{max} + E_{min}}{2} \qquad \text{... (1.2.7)}$$

Taking the ratio of equation (1.2.5) and above equation,

$$m = \frac{E_m}{E_c} = \frac{\frac{E_{max} - E_{min}}{2}}{\frac{E_{max} + E_{min}}{2}}$$

$$\therefore \quad \boxed{m = \frac{E_{max} - E_{min}}{E_{max} + E_{min}}} \qquad \text{... (1.2.8)}$$

This equation gives the technique of calculating modulation index from AM wave.

1.2.3 Frequency Spectrum and Bandwidth

The modulated carrier has new signals at different frequencies, called side frequencies or sidebands. They occur above and below the carrier frequency.

i.e.

$$f_{USB} = f_c + f_m$$

$$f_{LSB} = f_c - f_m$$

Here f_c is carrier frequency and

f_m is modulating signal frequency

f_{LSB} is lower sideband frequency

Consider the expression of AM wave given by equation (1.2.3), i.e.,

$$e_{AM} = (E_c + E_m \sin \omega_m t) \sin \omega_c t \qquad \text{... (1.2.9)}$$

We know that $m = \frac{E_m}{E_c}$ from equation (1.2.4). Hence we have $E_m = m E_c$. Putting this value of E_m in above equation we get,

$$\begin{aligned} e_{AM} &= (E_c + m E_c \sin \omega_m t) \sin \omega_c t \\ &= E_c (1 + m \sin \omega_m t) \sin \omega_c t \\ &= E_c \sin \omega_c t + m E_c \sin \omega_m t \sin \omega_c t \qquad \text{... (1.2.10)} \end{aligned}$$

We know that $\sin(A) \sin(B) = \frac{1}{2} \cos(A - B) - \frac{1}{2} \cos(A + B)$. Applying this result to last term in above equation we get,

$$\begin{aligned} e_{AM} &= E_c \sin \omega_c t + \frac{m E_c}{2} \cos(\omega_c - \omega_m) t \\ &\quad - \frac{m E_c}{2} \cos(\omega_c + \omega_m) t \qquad \text{... (1.2.11)} \end{aligned}$$

In the above equation, the first term represents unmodulated carrier, the second term represents lower sideband and last term represents upper sideband. Note that $\omega_c = 2\pi f_c$ and $\omega_m = 2\pi f_m$. Hence above equation can also be written as,

$$\begin{aligned} e_{AM} &= E_c \sin 2\pi f_c t + \frac{m E_c}{2} \cos 2\pi (f_c - f_m) t \\ &\quad - \frac{m E_c}{2} \cos 2\pi (f_c + f_m) t \qquad \text{... (1.2.12)} \\ &= E_c \sin 2\pi f_c t + \frac{m E_c}{2} \cos 2\pi f_{LSB} t + \frac{m E_c}{2} \cos 2\pi f_{USB} t \qquad \text{... (1.2.13)} \end{aligned}$$

From this equation we can prepare the frequency spectrum of AM wave as shown below in Fig. 1.2.3.

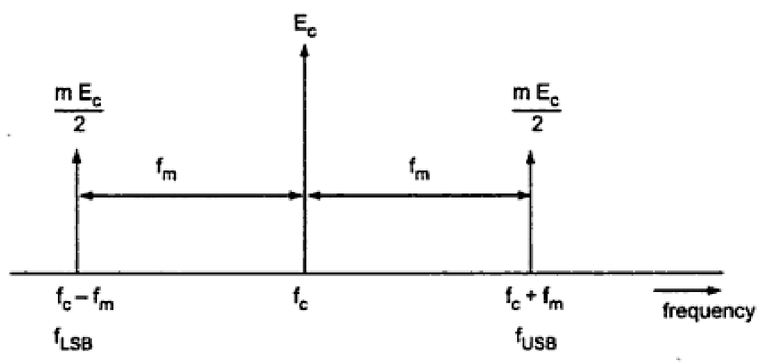


Fig. 1.2.3 Frequency domain representation of AM wave

This contains full carrier and both the sidebands, hence it is also called Double Sideband Full Carrier (DSBFC) system. We will be discussing this system, its modulation circuits and transmitters next, in this section.

We know that bandwidth of the signal can be obtained by taking the difference between highest and lowest frequencies. From above figure we can obtain bandwidth of AM wave as,

$$BW = f_{USB} - f_{LSB}$$
$$= (f_c + f_m) - (f_c - f_m)$$

$$\therefore \quad \boxed{BW = 2f_m} \quad \text{... (1.2.14)}$$

Thus bandwidth of AM signal is twice of the maximum frequency of modulating signal.

Example 1.2.1 : *Calculate the modulation index and percentage modulation if instantaneous voltages of modulating signal and carrier are $40 \sin \omega_m t$ and $50 \sin \omega_c t$, respectively.*

Solution : From the given instantaneous equation we have,

$$E_m = 40 \quad \text{and} \quad E_c = 50$$

Hence modulation index will be,

$$m = \frac{E_m}{E_c} = \frac{40}{50} = 0.8$$

or

$$\% \text{ modulation} = m \times 100$$
$$= 0.8 \times 100 = 80\%$$

Example 1.2.2 : *The tuned circuit of the oscillator in a simple AM transmitter employs a $40 \mu H$ coil and 12 nF capacitor. If the oscillator output is modulated by audio frequency of 5 kHz, what are the lower and upper sideband frequencies and the bandwidth required to transmit this AM wave ?*

Solution : The frequency of the LC oscillator is given as,

$$f_c = \frac{1}{2\pi\sqrt{LC}}$$
$$= \frac{1}{2\pi\sqrt{40 \times 10^{-6} \times 12 \times 10^{-9}}}$$
$$= 230 \text{ kHz}$$

The modulating frequency is $f_m = 5\,kHz$

$$\therefore \quad f_{USB} = f_c + f_m = 230 + 5 = 235\,kHz$$

and

$$f_{LSB} = f_c - f_m = 230 - 5 = 225\,kHz$$

We know that bandwidth of AM wave is,

$$BW = 2f_m$$
$$= 2 \times 5\,kHz = 10\,kHz$$

1.2.4 AM Power Distribution

May/ June - 2006

We know that AM signal has three components : Unmodulated carrier, lower sideband and upper sideband. Hence total power of AM wave is the sum of carrier power P_c and powers in the two sidebands P_{USB} and P_{LSB}. i.e.,

$$P_{Total} = P_c + P_{USB} + P_{LSB}$$
$$= \frac{E_{carr}^2}{R} + \frac{E_{LSB}^2}{R} + \frac{E_{USB}^2}{R} \quad \text{... (1.2.15)}$$

Here all the three voltages are rms values and R is characteristic impedence of antenna in which the power is dissipated. The carrier power is,

$$P_c = \frac{E_{carr}^2}{R} = \frac{\left(E_c / \sqrt{2}\right)^2}{R}$$

$$= \frac{E_c^2}{2R} \qquad \text{... (1.2.16)}$$

The power of upper and lower sidebands is same. i.e.,

$$P_{LSB} = P_{USB} = \frac{E_{SB}^2}{R} \qquad \text{Here } E_{SB} \text{ is rms voltage of sidebands.}$$

From equation (1.2.13) we know that the peak amplitude of both the sidebands is $\frac{m E_c}{2}$. Hence,

$$E_{SB} = \frac{m E_c / 2}{\sqrt{2}}$$

$$\therefore \qquad P_{LSB} = P_{USB} = \left(\frac{m E_c / 2}{\sqrt{2}}\right)^2 \times \frac{1}{R}$$

$$= \frac{m^2 E_c^2}{8R} \qquad \text{... (1.2.17)}$$

Hence the total power (equation 1.2.15) becomes,

$$P_{Total} = \frac{E_c^2}{2R} + \frac{m^2 E_c^2}{8R} + \frac{m^2 E_c^2}{8R}$$

$$= \frac{E_c^2}{2R}\left[1 + \frac{m^2}{4} + \frac{m^2}{4}\right]$$

$$\therefore \qquad \boxed{P_{Total} = P_c\left(1 + \frac{m^2}{2}\right)} \qquad \text{... (1.2.19)}$$

$$\frac{P_{Total}}{P_c} = 1 + \frac{m^2}{2} \qquad \text{... (1.2.20)}$$

This equation relates total power of AM wave to carrier power. Maximum value of modulation index, $m = 1$ to avoid distortion. At this value of modulation index, $P_{Total} = 1.5\, P_c$. From above equation we have,

$$\frac{m^2}{2} = \frac{P_{Total}}{P_c} - 1$$

$$\therefore \qquad m = \sqrt{2\left(\frac{P_{total}}{P_c} - 1\right)} \qquad \text{... (1.2.21)}$$

Example 1.2.3 : *An audio frequency signal $10 \sin 2\pi \times 500\, t$ is used to amplitude modulate a carrier of $50 \sin 2\pi \times 10^5\, t$. Calculate*

(i) Modulation index

(ii) Sideband frequencies

(iii) Amplitude of each sideband frequencies

(iv) Bandwidth required

(v) Total power delivered to the load of $600\,\Omega$.

Solution : (i) The given modulating signal is $e_m = 10 \sin 2\pi \times 500 t$. Hence, $E_m = 10$. The given carrier signal is $e_c = 50 \sin 2\pi \times 10^5\, t$, hence, $E_c = 50$. Therefore modulation index will be,

$$m = \frac{E_m}{E_c} = \frac{10}{50} = 0.2 \quad \text{or} \quad 20\%$$

(ii) From the given equations,

$$\omega_m = 2\pi \times 500, \qquad \text{Hence } f_m = 500\, Hz$$

And $$\omega_c = 2\pi \times 10^5, \qquad \text{Hence } f_c = 10^5\ Hz \text{ or } 100 \text{ kHz}$$

We know that $$f_{USB} = f_c + f_m = 100\, kHz + 500\, Hz = 100.5\, kHz$$

and $$P_{LSB} = f_c - f_m = 100\, kHz - 500\, Hz = 99.5\, kHz.$$

(iii) From equation (1.2.13) we know that the amplitudes of upper and lower sidebands is given as,

$$\text{Amplitude of upper and lower sidebands} = \frac{m\, E_c}{2} = \frac{0.2 \times 50}{2} = 5V$$

(iv) Bandwidth of AM wave is given by equation(1.2.10) as,

$$BW \text{ of } AM = 2 f_m = 2 \times 500\, Hz = 1\, kHz$$

(v) Total power delivered to the load is given by equation (1.2.18) as

$$P_{total} = \frac{E_c^2}{2R}\left(1 + \frac{m^2}{2}\right) = \frac{50^2}{2 \times 600}\left(1 + \frac{(0.2)^2}{2}\right)$$

$$\therefore \qquad = 2.125 \text{ watts}$$

Example 1.2.4 : *A 400 W carrier is modulated to a depth of 80% calculate the total power in the modulated wave.*

Solution : Here carrier power $P_c = 400\, W$ and $m = 0.8$.

From equation (1.2.19) total power is,

$$P_{total} = P_c\left(1+\frac{m^2}{2}\right) = 400\left(1+\frac{(0.8)^2}{2}\right)$$

$$= 528\,W$$

Example 1.2.5 : *A broadcast transmitter radiates 20 kW when the modulation percentage is 75. Calculate carrier power and power of each sideband.*

Solution : Here total power $P_{total} = 20{,}000\,W$ and $m = 0.75$

From equation (1.2.19) we have $P_{total} = P_c\left(1+\frac{m^2}{2}\right)$

$$\therefore \quad 20{,}000 = P_c\left(1+\frac{(0.75)^2}{2}\right)$$

$$\therefore \quad P_c = 15.6\,kW$$

We know that $$P_{total} = P_c\left(1+\frac{m^2}{2}\right) = P_c + P_c\,\frac{m^2}{2}$$

The second term in above equation represents total sideband power. Hence power of one sideband will be,

$$P_{SB} = \left(P_c\,\frac{m^2}{2}\right)\times\frac{1}{2}$$

$$= 15.6\times\frac{(0.75)^2}{2}\times\frac{1}{2}$$

$$= 2.2\,kW$$

Thus $$P_{USB} = P_{LSB} = 2.2\,kW$$

2.1 Angle Modulation

2.1.1 Definition

We know that amplitude, frequency or phase of the carrier can be varied by the modulating signal. Amplitude is varied in AM. *When frequency or phase of the carrier is varied by the modulating signal, then it is called angle modulation.* There are two types of angle modulation.

1. Frequency Modulation : When frequency of the carrier varies as per amplitude variations of modulating signal, then it is called Frequency Modulation (FM). Amplitude of the modulated carrier remains constant.

2. Phase Modulation : When phase of the carrier varies as per amplitude variations of modulating signal, then it is called Phase Modulation (PM). Amplitude of the modulated carrier remains constant.

The angle modulated wave is mathematically expressed as,

$$e(t) = E_c \sin[\omega_c t + \theta(t)]$$

Here $e(t)$ is angle modulated wave

E_c is peak amplitude of the carrier

ω_c carrier frequency

$\theta(t)$ instantaneous phase deviation

The phase deviation takes place in FM as well as PM. Hence phase is direct function of modulating signal. i.e.,

$$\theta(t) \propto e_m(t)$$

Here $e_m(t)$ is the modulating signal.

2.1.3 FM and PM Waveforms

Fig. 2.1.1 shows the waveforms of FM and PM.

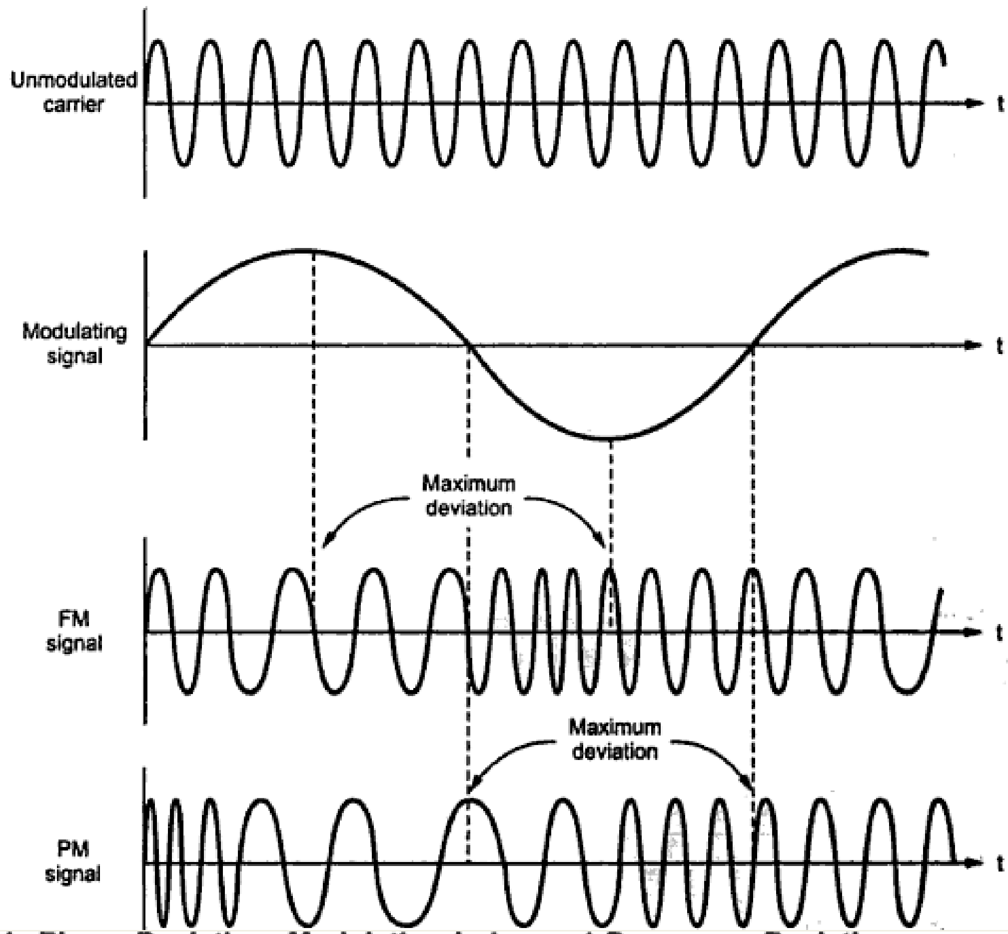
In this figure following observations can be noted :

(i) For FM signal, the maximum frequency deviation takes place when modulating signal is at positive and negative peaks.

(ii) For PM signal the maximum frequency deviation takes place near zero crossings of the modulating signal.

(iii) Both FM and PM waveforms are identical except the phase shift.

(iv) From modulated waveform it is difficult to know, whether the modulation is FM or PM.



2.1.4 Phase Deviation, Modulation Index and Frequency Deviation

The FM signal, in general is expressed as,

$$e_{FM}(t) = E_c \sin[\omega_c t + m \sin \omega_m t] \quad \text{... (2.1.12)}$$

And the PM signal, in general is expressed as,

$$e_{PM}(t) = E_c \sin[\omega_c t + m \cos \omega_m (t)] \quad \text{... (2.1.13)}$$

In both the above equations, the term 'm' is called *modulation index*. Note that the term $m \sin \omega_m t$ in equation 2.1.12 and $m \cos \omega_m t$ in equation 2.1.13 indicates instantaneous phase deviation $\theta(t)$. Hence 'm' also indicates *maximum phase deviation*. In other words, modulation index can also be defined as maximum phase deviation.

Modulation index for PM :

Comparing equation 2.1.13 and equation 2.1.11, we find that,

$$\boxed{\textbf{Modulation index in PM : } m = k E_m \text{ rad}} \quad \text{... (2.1.14)}$$

Thus modulation index of PM signal is directly proportional to peak modulating voltage. And it's unit is radians.

Modulation index for FM :

Comparing equation 2.1.12 and equation 2.1.10 we find that,

$$m = \frac{k_1 E_m}{\omega_m} \qquad \text{... (2.1.15)}$$

Thus modulation index of FM is directly proportional to peak modulating voltage, but inversely proportional to modulating signal frequency.

$$\text{Since} \quad \omega_m = 2\pi f_m \text{ above equation becomes,}$$

$$m = \frac{k_1 E_m}{2\pi f_m}$$

Here $\frac{k_1 E_m}{2\pi}$ is called *frequency deviation.* It is denoted by δ and its unit is Hz, i.e.,

$$\boxed{\textbf{Modulation index in FM :} \quad m = \frac{\delta}{f_m} = \frac{\text{Maximum frequency deviation}}{\text{Modulating frequency}}} \qquad \text{... (2.1.16)}$$

Thus modulation index of FM is unitless ratio. From above equation and equation 2.1.14, note that the modulation index is differently defined for FM and PM signals.

Percentage modulation :

For angle modulation, the percentage modulation is given as the ratio of actual frequency deviation to maximum allowable frequency deviation. i.e.,

$$\boxed{\%\text{ modulation} = \frac{\text{Actual frequency deviation}}{\text{Maximum allowable frequency deviation}}} \qquad \text{... (2.1.17)}$$

2.1.5 Frequency Spectrum of Angle Modulated Waves

We know that AM contains only two sidebands per modulating frequency. But angle modulated signal contains large number of sidebands depending upon the modulation index. Since FM and PM have identical modulated waveforms, their frequency content is same. Consider the PM equation for spectrum analysis,

$$e(t) = E_c \sin\left[\omega_c t + m\cos\omega_m t\right]$$

Using Bessel functions, this equation can be expanded as,

$$\begin{aligned} e(t) = E_c \{ & J_0 \sin\omega_c t \\ & + J_1\left[\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t\right] \\ & + J_2\left[\sin(\omega_c + 2\omega_m)t + \sin(\omega_c - 2\omega_m)t\right] \\ & + J_3\left[\sin(\omega_c + 3\omega_m)t + \sin(\omega_c - 3\omega_m)t\right] \\ & + J_4\left[\sin(\omega_c + 4\omega_m)t - \sin(\omega_c - 4\omega_m)t\right] + \ldots \} \end{aligned}$$

Here $J_0, J_1, J_2 \ldots$ are the Bessel functions. The values of Bessel functions depend upon modulation index m. They are listed in Table 2.1.1

x	n or Order																
(m)	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}	J_{15}	J_{16}
0.00	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0.25	0.98	0.12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.94	0.24	0.03	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1.0	0.77	0.44	0.11	0.02	—	—	—	—	—	—	—	—	—	—	—	—	—
1.5	0.51	0.56	0.23	0.06	0.01	—	—	—	—	—	—	—	—	—	—	—	—
2.0	0.22	0.58	0.35	0.13	0.03	—	—	—	—	—	—	—	—	—	—	—	—
2.5	−0.05	0.50	0.45	0.22	0.07	0.02	—	—	—	—	—	—	—	—	—	—	—
3.0	−0.26	0.34	0.49	0.31	0.13	0.04	0.01	—	—	—	—	—	—	—	—	—	—
4.0	−0.40	−0.07	0.36	0.43	0.28	0.13	0.05	0.02	—	—	—	—	—	—	—	—	—
5.0	−0.18	−0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02	—	—	—	—	—	—	—	—
6.0	0.15	−0.28	−0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	—	—	—	—	—	—	—
7.0	0.30	0.00	−0.30	−0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02	—	—	—	—	—	—
8.0	0.17	0.23	−0.11	−0.29	−0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03	0.01	—	—	—	—
9.0	−0.09	0.24	0.14	−0.18	−0.27	−0.06	0.20	0.33	0.30	0.21	0.12	0.06	0.03		—	—	—
10.0	−0.25	0.04	0.25	0.06	−0.22	−0.23	−0.01	0.22	0.31	0.29	0.20	0.12	0.06	0.03	0.01	—	—
12.0	0.05	−0.22	−0.08	0.20	0.18	−0.07	−0.24	−0.17	0.05	0.23	0.30	0.27	0.20	0.12	0.07	0.03	0.01
15.0	−0.01	0.21	0.04	−0.19	−0.12	0.13	0.21	0.03	−0.17	−0.22	−0.09	0.10	0.24	0.28	0.25	0.18	0.12

It is clear from the above discussion that, angle modulated signal has infinite number of sidebands as well as carrier in the output. The sidebands are separated from the carrier by f_m, $2f_m$, $3f_m$..... etc. The frequency separation between successive sidebands is f_m. All the sidebands are symmetric around carrier frequency. The amplitudes of the sidebands are $E_c J_0$, $E_c J_1$, $E_c J_2$, $E_c J_3$, $E_c J_4$, ... and so on.

2.1.6 Bandwidth Requirement

The bandwidth requirement of angle modulated waveforms can be obtained depending upon modulation index. The modulation index can be classified as low (less than 1), medium (1 to 10) and high (greater than 10). The low index systems are called *narrowband FM*. For such systems the frequency spectrum resembles AM. Hence minimum bandwidth is given as,

$$BW = 2 f_m \ Hz \qquad \text{... (2.1.19)}$$

For high index modulation, the minimum bandwidth is given as,

$$BW = 2\delta \qquad \text{... (2.1.20)}$$

The bandwidth can also be obtained using bessel table. i.e.,

$$BW = 2 n f_m \qquad \text{... (2.1.21)}$$

Here 'n' is the number of significant sidebands obtained from bessel table.

Carson's rule :

Carson's rule gives approximate minimum bandwidth of angle modulated signal as,

$$\boxed{BW = 2\left[\delta + f_{m(\max)}\right] Hz} \qquad \text{... (2.1.22)}$$

Here $f_{m(\max)}$ is the maximum modulating frequency. As per Carson's rule, the bandwidth accomodates almost 98% of the total transmitted power.

UNIT II DIGITAL COMMUNICATION

INTRODUCTION:

2.2 Shannon Limit for Information Capacity

Definition of information capacity : It is an ability of the system to carry number of independent symbols in a given unit of time. The capacity is expressed in bits per second.

Shannon's limit for information capacity

The capacity of a white bandlimited gaussian channel is given as,

$$C = B \log_2\left(1+\frac{S}{N}\right) \text{ bits/sec} \qquad \text{...(2.2.1)}$$

Here C is the channel capacity

B is the channel bandwidth

$\frac{S}{N}$ is the signal to noise power ratio.

2.3 Digital Amplitude Modulation or Amplitude Shift Keying

The amplitude shift keying is also called on-off keying (OOK). This is the simplest digital modulation technique. The binary input data is converted to unipolar NRZ signal. A product modulator takes this NRZ signal and carrier signal. The output of the product modulator is the ASK signal, which can be expressed mathematically as,

$$v(t) = d \sin(2\pi f_c t) \qquad \text{... (2.3.1)}$$

Here f_c is the carrier frequency

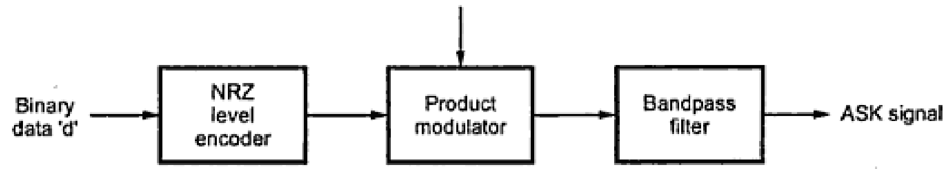
and d is the data bit, which is either 1 or 0.

Fig. 2.3.1 (a) shows the block diagram of the ASK modulator. The binary data sequence 'd' is given to the NRZ level encoder. This NRZ level encoder converts the input binary sequence to the signal suitable for product modulator. The product modulator also accepts a sinusoidal carrier of frequency f_c. The output of the product modulator is passed through a bandpass filter for bandwidth limiting. The output of the bandpass filter is the ASK signal. This signal and other waveforms are shown in Fig. 2.3.1 (b). Observe that the ASK signal has on-off nature. In equation 2.3.1 when $d=0$, $v(t)=0$; i.e. no ASK signal. And when $d=1$, $d=\sin(2\pi f_c t)$. The ASK is very sensitive to noise. It is used for very low bit rates less than around 100 bps. The only advantage of ASK is that it is very simple to implement.

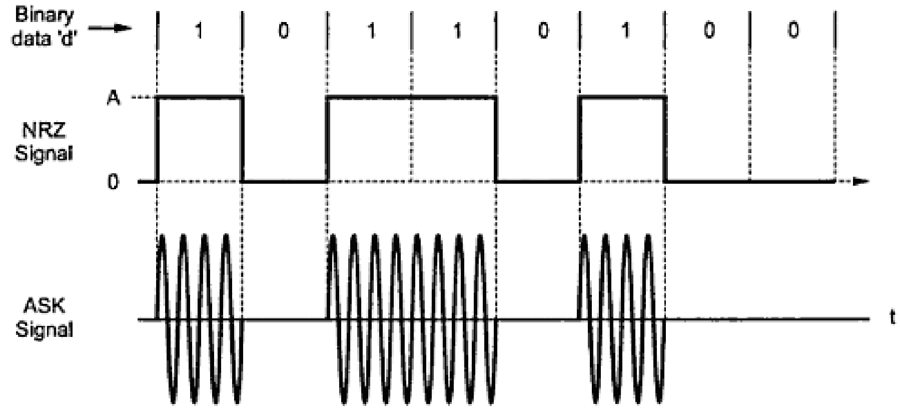
Baud rate

For ASK, the ASK waveform is changed at the bit rate. Hence Baud rate is given as,

$$\text{Baud rate} = f_b \qquad \text{... (2.3.2)}$$



(a) Block diagram of ASK modulator



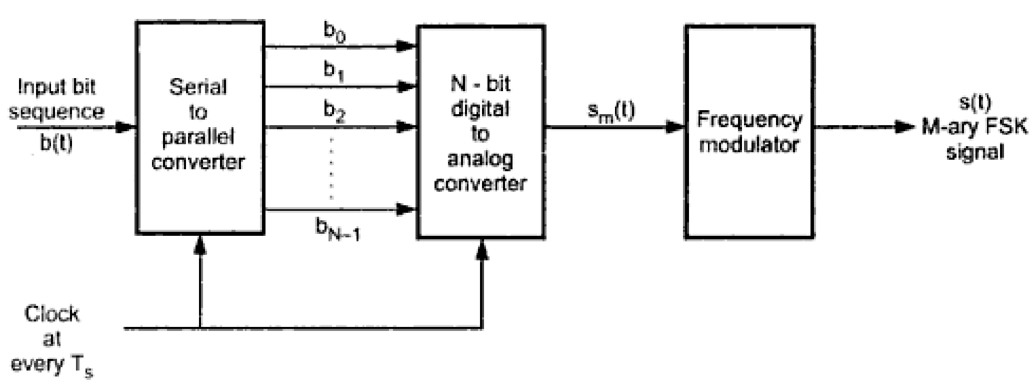
(b) Waveforms

Fig. 2.3.1 Amplitude shift keying (ASK)

FREQUENCY SHIFT KEYING:

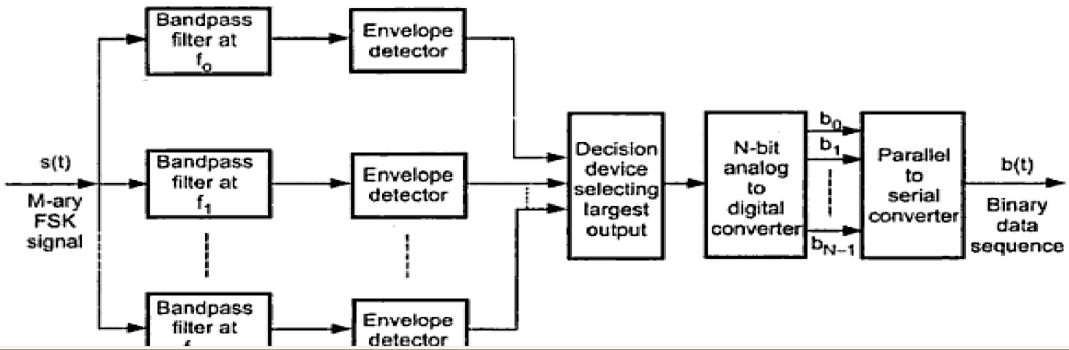
3.8.1.1 Transmitter

Fig. 3.8.1 shows the M-ary FSK transmitter. The 'N' successive bits are presented in parallel to digital to analog converter. These 'N' bits forms a symbol at the output of digital to analog converter. There will be total $2^N = M$ possible symbols. The symbol is presented every $T_s = NT_b$ period. The output of digital to analog converter is given to a frequency modulator. Thus depending upon the value of symbol, the frequency modulator generates the output frequency. For every symbol, the frequency modulator produces different frequency output. This particular frequency signal remains at the output for one symbol duration. Thus for 'M' symbols, there are 'M' frequency signals at the output of modulator. Thus the transmitted frequencies are $f_0, f_1, f_2, \ldots f_{M-1}$ depending upon the input symbol to the modulator.



3.8.1.2 Receiver

Fig. 3.8.2 shows block diagram of M-ary FSK receiver. It is the extension of BFSK receiver of Fig. 3.8.1. The M-ary FSM signal is given to the set of 'M' bandpass filters. The center frequencies of those filters are $f_0, f_1, f_2, \ldots f_{M-1}$. These filters pass their particular frequency and alternate others. The envelope detectors outputs are applied to a decision device. The decision device produces its output depending upon the highest input. Depending upon the particular symbol, only one envelope detector will have higher output. The outputs of other detectors will be very low. The output of the decision device is given to 'N' bit analog to digital converter. The analog to digital converter output is the 'N' bit symbol in parallel. These bits are then converted to serial bit stream by parallel to serial converter. In some cases the bits appear in parallel. Then there is no need to use serial to parallel and parallel to serial converters.



3.8.2 Power Spectral Density and Bandwidth of M-ary FSK

We know that for M symbol $f_0, f_1, f_2 \ldots f_{m-1}$ frequencies are used for transmission. The probability of error is minimized by selecting those frequencies such that transmitted signals are mutually orthogonal. If those frequencies are selected as successive even harmonics of symbol frequency f_s, then transmitted signals will be orthogonal.

Let's say that the lowest carrier frequency f_0 is the k^{th} harmonic of symbol frequency i.e.,

$$f_0 = kf_s \qquad \ldots (3.8.1)$$

Then the other frequencies will be,

$$f_1 = (k+2) f_s, f_2 = (k+4) f_s \ \ldots \text{etc} \qquad \ldots (3.8.2)$$

Thus every carrier frequency is separated by $2f_s$ from its nearest carriers. Fig.3.7.2 shows the power spectral density of BFSK (for two symbol FSK). In this plot the two symbol frequencies f_L and f_H are separated by $2f_s$ (Here $f_s = f_b$ for BFSK). The same principle of BFSK is extended to M-ary FSK. That is M-carriers are added with separation of $2f_s$ between the carrriers (Note here that f_s is symbol frequency and not f_b). Therefore power spectral density for M-ary FSK will be simply extension of BFSK. Fig. 3.8.3 shows the power spectral density of M-ary FSK.

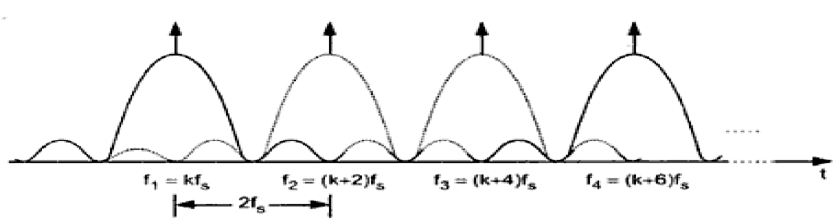


Fig. 3.8.3 Power spectral density M-ary FSK

Bandwidth of M-ary FSK :

From Fig. 3.8.3 it is clear that the width of one main lobe is $2f_s$. If there are M-symbols, then power spectral density spectrum will have M lobes. Therefore bandwidth of the system for M-symbols will be

$$BW = M \times (2f_s)$$

$$= 2Mf_s \qquad \dots (3.8.3)$$

We know that $2^N = M$ and $f_s = \frac{f_b}{N}$ we can write the above equations,

$$BW = 2 \cdot 2^N \cdot \frac{f_b}{N} \qquad \dots (3.8.4)$$

$$= \frac{2^{N+1} f_b}{N} \qquad \dots (3.8.5)$$

4.2 Binary Phase Shift Keying (BPSK)

4.2.1 Principle of BPSK

- In binary phase shift keying (BPSK), binary symbol '1' and '0' modulate the phase of the carrier. Let the carrier be,

$$s(t) = A \cos(2\pi f_0 t) \qquad \dots (4.2.1)$$

'A' represents peak value of sinusoidal carrier. In the standard 1Ω load register, the power dissipated will be,

$$P = \frac{1}{2} A^2$$

$$\therefore \quad A = \sqrt{2P} \qquad \dots (4.2.2)$$

- When the symbol is changed, then the phase of the carrier is changed by 180 degrees (π radians).
- Consider for example,

$$\text{Symbol '1'} \Rightarrow s_1(t) = \sqrt{2P} \cos(2\pi f_0 t) \qquad \dots (4.2.3)$$

if next symbol is '0' then,

$$\text{Symbol '0'} \Rightarrow s_2(t) = \sqrt{2P} \cos(2\pi f_0 t + \pi) \qquad \dots (4.2.4)$$

Since $\cos(\theta + \pi) = -\cos\theta$, we can write above equation as,

$$s_2(t) = -\sqrt{2P} \cos(2\pi f_0 t) \qquad \dots (4.2.5)$$

With the above equation we can define BPSK signal combinely as,

$$\boxed{s(t) = b(t)\sqrt{2P} \cos(2\pi f_0 t)} \qquad \dots (4.2.6)$$

Here $b(t) = +1$ when binary '1' is to be transmitted

$= -1$ when binary '0' is to be transmitted

4.2.2 Graphical Representation of BPSK Signal

Fig. 4.2.1 shows binary signal and its equivalent signal $b(t)$.

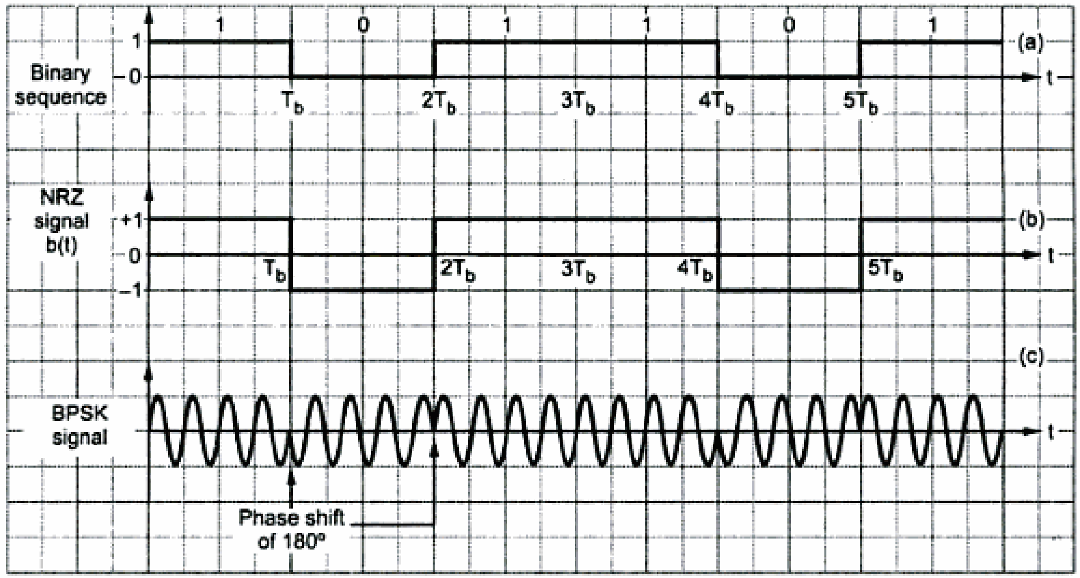


Fig. 4.2.1 (a) Binary sequence
(b) Its equivalent bipolar signal $b(t)$
(c) BPSK signal

As can be seen from Fig. 4.2.1 (b), the signal $b(t)$ is NRZ bipolar signal. This signal directly modulates carrier $cos(2\pi f_0 t)$.

4.2.3 Generation and Reception of BPSK Signal

4.2.3.1 Generation of BPSK Signal

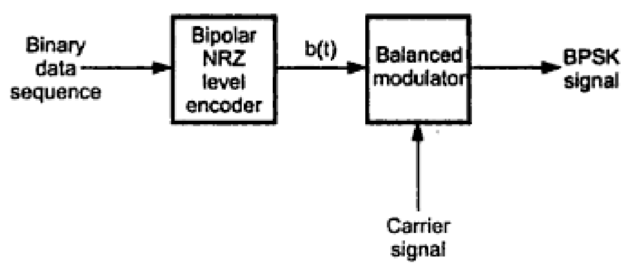


Fig. 4.2.2 BPSK generation scheme

- The BPSK signal can be generated by applying carrier signal to the balanced modulator.
- The baseband signal $b(t)$ is applied as a modulating signal to the balanced modulator. Fig. 4.2.2 shows the block diagram of BPSK signal generator.
- The NRZ level encoder converts the binary data sequence into bipolar NRZ signal.

4.2.3.2 Reception of BPSK Signal

Fig. 4.2.3 shows the block diagram of the scheme to recover baseband signal from BPSK signal. The transmitted BPSK signal is,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t)$$

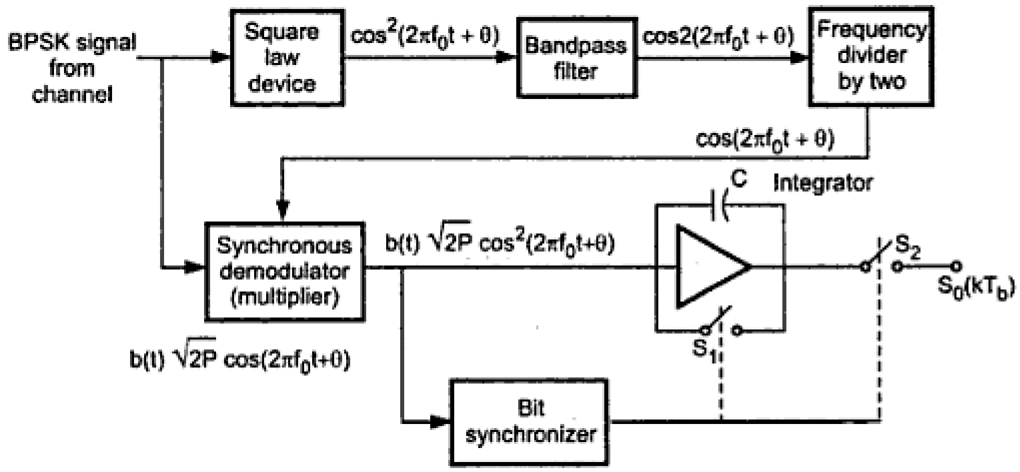


Fig. 4.2.3 Reception BPSK scheme

Operation of the receiver

1) **Phase shift in received signal** : This signal undergoes the phase change depending upon the time delay from transmitter to receiver. This phase change is normally fixed phase shift in the transmitted signal. Let the phase shift be θ. Therefore the signal at the input of the receiver is,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t + \theta) \qquad \text{... (4.2.7)}$$

2) **Square law device** : Now from this received signal, a carrier is separated since this is coherent detection. As shown in the figure, the received signal is passed through a square law device. At the output of the square law device the signal will be,

$$\cos^2(2\pi f_0 t + \theta)$$

Note here that we have neglected the amplitude, because we are only interested in the carrier of the signal.

We know that,

$$\cos^2\theta = \frac{1 + \cos 2\theta}{2}$$

$$\therefore \quad \cos^2(2\pi f_0 t + \theta) = \frac{1 + \cos 2(2\pi f_0 t + \theta)}{2}$$

4.2.4 Spectrum of BPSK Signals

Step 1 : Fourier transform of basic NRZ pulse.

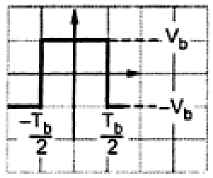


Fig. 4.2.4 NRZ pulse

We know that the waveform $b(t)$ is NRZ bipolar waveform. In this waveform there are rectangular pulses of amplitude $\pm V_b$. If we say that each pulse is $\pm \frac{T_b}{2}$ around its center as shown in Fig. 4.2.4. then it becomes easy to find fourier transform of such pulse. The fourier transform of this type of pulse is given as,

$$X(f) = V_b T_b \frac{\sin(\pi f T_b)}{(\pi f T_b)} \quad \text{By standard relations} \qquad \text{... (4.2.10)}$$

Step 2 : PSD of NRZ pulse.

For large number of such positive and negative pulses the power spectral density $S(f)$ is given as

$$S(f) = \frac{\overline{|X(f)|^2}}{T_s} \qquad \text{... (4.2.11)}$$

Here $\overline{X(f)}$ denotes average value of $X(f)$ due to all the pulses in $b(t)$. And T_s is symbol duration. Putting value of $X(f)$ from equation 4.2.10 in equation 4.2.11 we get,

Plot of PSD

- Equation 4.2.12 gives power spectral density of the NRZ waveform. For one rectangular pulse, the shape of $S(f)$ will be a sinc pulse as given by equation 4.2.12. Fig. 4.2.5 shows the plot of magnitude of $S(f)$.

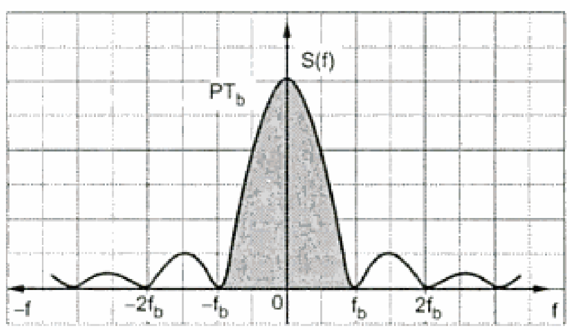


Fig. 4.2.5 Plot of power spectral density of NRZ baseband signal

Above figure shows that the main lobe ranges from $-f_b$ to $+f_b$. Here $f_b = \frac{1}{T_b}$.

Since we have taken $\pm V_b = \pm\sqrt{P}$ in equation 4.2.12, the peak value of the main lobe is PT_b.

- Now let us consider the power spectral density of BPSK signal given by equation 4.2.13. Fig. 4.2.6 shows the plot of this equation. The figure thus clearly shows that there are two lobes ; one at f_0 and other at $-f_0$. The same spectrum of Fig. 4.2.5 is placed at $+f_0$ and $-f_0$. But the amplitudes of main lobes are $\frac{PT_b}{2}$ in Fig. 4.2.6.

4.2.6 Bandwidth of BPSK Signal

The spectrum of the BPSK signal is centered around the carrier frequency f_0.

If $f_b = \frac{1}{T_b}$, then for BPSK the maximum frequency in the baseband signal will be

f_b see Fig. 4.2.6. In this figure the main lobe is centered around carrier frequency f_0 and extends from $f_0 - f_b$ to $f_0 + f_b$. Therefore Bandwidth of BPSK signal is,

$$BW = \text{Highest frequency} - \text{Lowest frequency in the main lobe}$$

$$= f_0 + f_b - (f_0 - f_b)$$

$$\therefore \qquad \boxed{BW = 2 f_b} \qquad \text{... (4.2.21)}$$

Thus the minimum bandwidth of BPSK signal is equal to twice of the highest frequency contained in baseband signal.

4.4 Quadrature Phase Shift Keying (QPSK)

Principle

- In communication systems we know that there are two main resources, i.e. transmission power and the channel bandwidth. The channel bandwidth depends upon the bit rate or signalling rate f_b. In digital bandpass transmission, a carrier is used for transmission. This carrier is transmitted over a channel.
- If two or more bits are combined in some symbols, then the signalling rate is reduced. Therefore the frequency of the carrier required is also reduced. This reduces the transmission channel bandwidth. Thus because of grouping of bits in symbols, the transmission channel bandwidth is reduced.
- In quadrature phase shift keying, two successive bits in the data sequence are grouped together. This reduces the bits rate of signalling rate (i.e. f_b) and hence reduces the bandwidth of the channel.
- In BPSK we know that when symbol changes the level, the phase of the carrier is changed by 180°. Since there were only two symbols in BPSK, the phase shift occurs in two levels only.
- In QPSK two successive bits are combined. This combination of two bits forms four distinct symbols. When the symbol is changed to next symbol the

Since $b_o(t)$ and $b_e(t)$ cannot change at the same time, the phase change in QPSK signal will be maximum $\pi/2$. This is clear from Fig. 4.4.3.

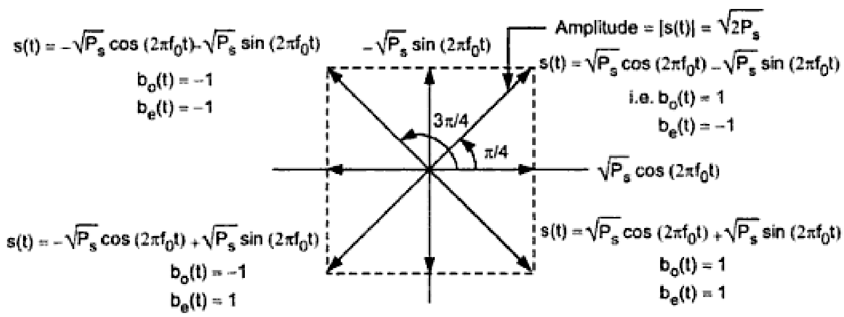


Fig. 4.4.3 Phasor diagram of QPSK signal

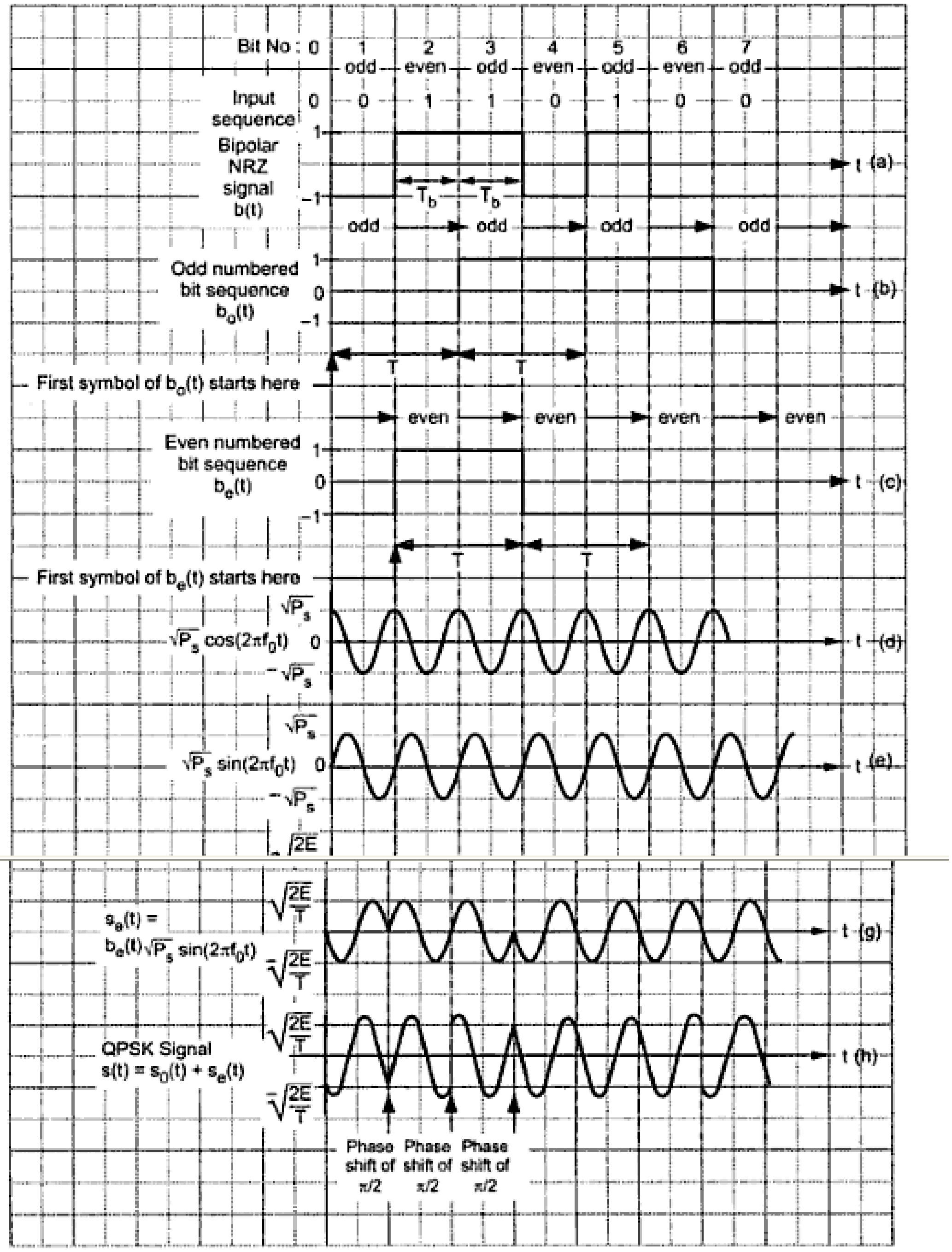


Fig. 4.4.2 QPSK waveforms (a) Input sequence and its NRZ waveform (b) Odd numbered bit sequence and its NRZ waveform (c) Even numbered bit sequence and its NRZ waveform (d) Basis function $\phi_1(t)$ (e) Basis function $\phi_2(t)$ (f) Binary PSK waveform for odd numbered channel (g) Binary PSK waveform for even numbered channel (h) Final QPSK waveform representing equation

4.4.1.3 The QPSK Receiver

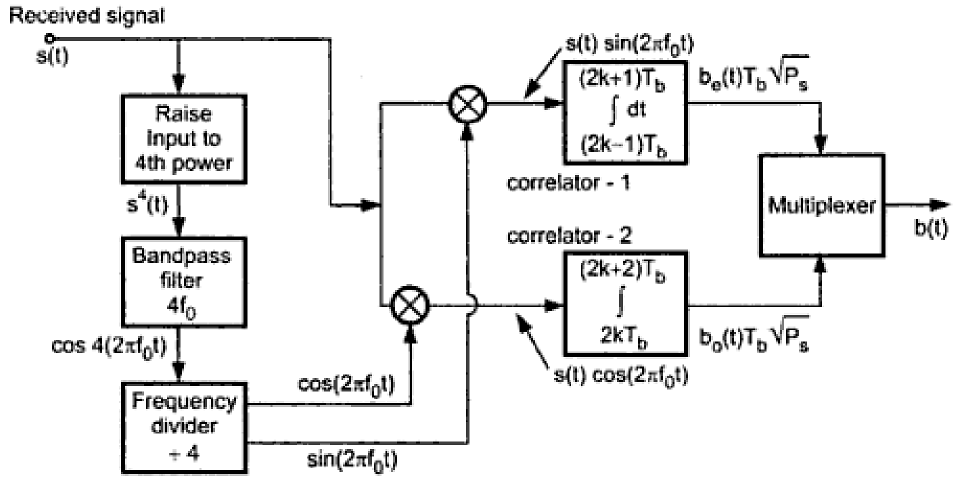


Fig. 4.4.4 QPSK receiver

Fig. 4.4.4 shows the QPSK receiver. This is synchronous reception. Therefore coherent carrier is to be recovered from the received signal $s(t)$.

Operation

Step 1 : Isolation of carrier

The received signal $s(t)$ is first raised to its 4^{th} power, i.e. $s^4(t)$. Then it is passed through a bandpass filter centered around $4f_0$. The output of the bandpass filter is a coherent carrier of frequency $4f_0$. This is divided by 4 and it gives two coherent quadrature carriers $\cos(2\pi f_0 t)$ and $\sin(2\pi f_0 (t)$.

Step 2 : Synchronous detection

These coherent carriers are applied to two synchronous demodulators. These synchronous demodulators consist of multiplier and an integrator.

Step 3 : Integration over two bits interval

The incoming signal is applied to both the multipliers. The integrator integrates the product signal over two bit interval (i.e. $T_s = 2T_b$).

Step 4 : Sampling and multiplexing odd and even bit sequences

At the end of this period, the output of integrator is sampled. The outputs of the two integrators are sampled at the offset of one bit period, T_b. Hence the output of

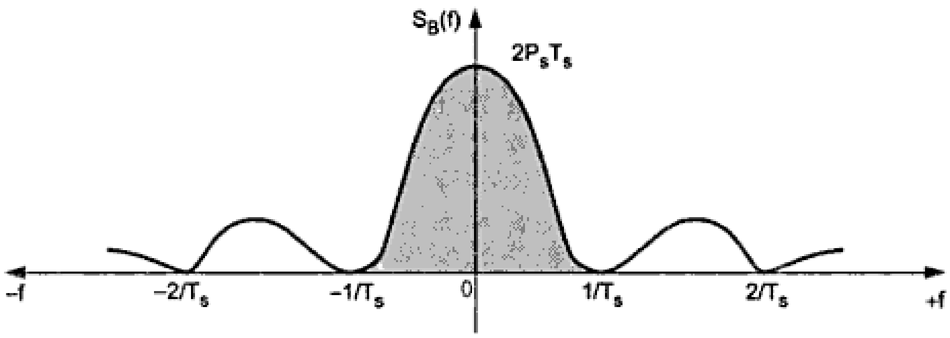


Fig. 4.4.7 Plot of power spectral density of QPSK signal

$$BW = \text{Highest frequency} - \text{Lowest frequency in main lobe}$$

$$= \frac{1}{T_s} - \left(-\frac{1}{T_s}\right) \text{ since carrier frequency } f_0 \text{ cancels out}$$

$$= \frac{2}{T_s}$$

We know that $T_s = 2T_b$

$$\therefore \qquad BW = \frac{2}{2T_b} = \frac{1}{T_b} = f_b$$

which is same as we obtained in equation 4.4.24.

4.4.5 Advantages of QPSK

QPSK has some definite advantages and disadvantages as compared to BPSK and DPSK.

Advantages :

1) For the same bit error rate, the bandwidth required by QPSK is reduced to half as compared to BPSK.
2) Because of reduced bandwidth, the information transmission rate of QPSK is higher.
3) Variation in OQPSK amplitude is not much. Hence carrier power almost remains constant.

2.9.1 Bandwidth Efficiency (Information Density)

Definition : It is the ratio of transmission bit rate to minimum required bandwidth.

i.e.,

$$\text{BW efficiency} = \frac{\text{Transmission rate(Bits / sec)}}{\text{Minimum bandwidth (cycles / sec)}}$$

$$= \frac{\text{Transmission rate}}{\text{Minimum bandwidth}} \text{ bits/cycle} \qquad ...(2.9.1)$$

- When bandwidth efficiency is normalized to 1-Hz bandwidth, it gives number of bits that can be propagated per hertz of bandwidth.
- Bandwidth efficiency is used to compare the performance of digital modulation techniques.

2.10 Carrier Synchronization (Carrier Recovery)

The carrier synchronization is required in coherent detection methods to generate a coherent reference at the receiver. In this method the data bearing signal is modulated on the carrier in such a way that the power spectrum of the modulated carrier signal contains a discrete component at the carrier frequency. That is the fourier transform of the modulated signal contains one component at f_c also. Then the phase locked loop can be used to track this component f_c. The output frequency of phase locked loop is thus locked to the carrier frequency f_c in the transmitted signal. This output frequency of phase locked loop is used as a coherent reference signal for detection in the receiver.

2.10.1 Carrier Synchronization using Mth Power Loop

Fig. 2.10.1 shows the block diagram of carrier recovery or carrier synchronization circuit.

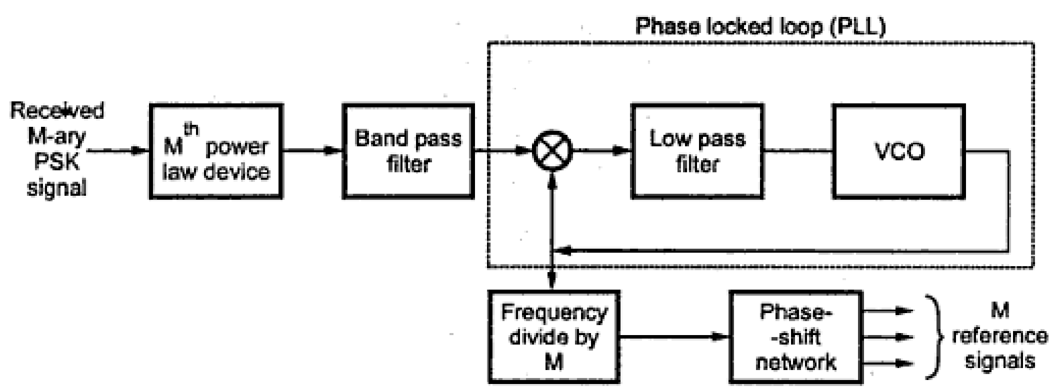


Fig. 2.10.1 Block diagram of Mth power loop

Fig. 2.10.1 shows the block diagram of carrier recovery circuit for M-ary PSK. This circuit is called the M^{th} power loop. When M = 2, then it is called squaring loop. When M=2, the M-ary PSK is then called as binary PSK. As shown in diagram, the input signal is first raised to the M^{th} power by the M^{th} power law device. Then the signal is passed through a bandpass filter. The bandpass filter is tuned to the carrier frequency f_c. The phase locked loop consists of a phase detector, low-pass filter and VCO. The phase locked loop tracks the carrier frequency. Then the output of a voltage controlled oscillator (VCO) is the carrier frequency. The output frequency of VCO is divided by M. This is done because the M^{th} power of the input signal multiplies carrier frequency by M. The phase shift network then separates 'M' reference signals for the 'M' correlation receivers. In this technique the power of the input signal is raised to some power say 'M'. Let us say M = 2, then the input signal is squared. Because of this, the sign of the recovered carrier is always independent of sign of the input signal carrier since it is squared. Therefore there can be 180° error in the output.

2.10.2 Costas Loop for Carrier Synchronization May / June - 2006

This is the alternative method for carrier synchronization. This is used for binary phase shift keying. The block diagram is shown in Fig. 2.10.2.

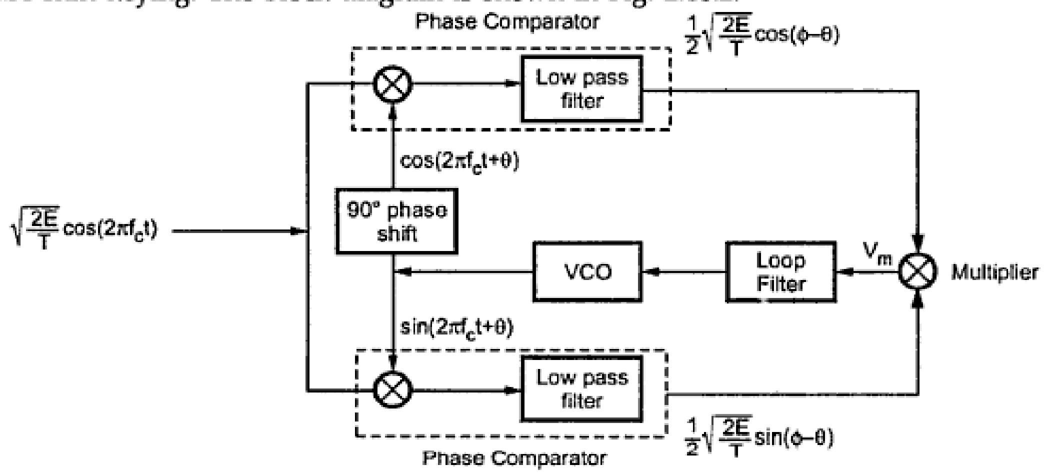


Fig. 2.10.2 The costas loop

As shown in Fig. 2.10.2 there are two phase locked loops. They have a common VCO and separate phase comparators. Let us assume that the VCO operates at the carrier frequency f_c with arbitrary phase angle θ. The BPSK signal is supplied to both the phase comparators. The low-pass filters remove the double frequency terms generated in the phase comparators and generate,

$\frac{1}{2}\sqrt{\frac{2E}{T}}\cos(\phi-\theta)$ and $\frac{1}{2}\sqrt{\frac{2E}{T}}\sin(\phi-\theta)$. The multiplier output is given as,

$$V_m = \frac{1}{4}\times\frac{2E}{T}\sin(\phi-\theta)\cos(\phi-\theta) \quad \text{... (2.10.1)}$$

$$= \frac{E}{2T}\cdot\frac{1}{2}\sin 2(\phi-\theta) \quad \text{..(2.10.2)}$$

$$= \frac{E}{4T}\sin 2(\phi-\theta) \quad \text{... (2.10.3)}$$

The power 'P' of the signal over the period T is given by,

$$P = \frac{E}{T}$$

Therefore equation (2.10.3) can be written as,

$$V_m = \frac{P}{4}\sin 2(\phi-\theta) \quad \text{... (2.10.4)}$$

If there is some difference between the VCO frequency and the input carrier frequency then the phase difference $(\phi-\theta)$ is changed proportionally. The change in $(\phi-\theta)$ causes V_m to increase or decrease VCO frequency such that synchronization is maintained.

2.5 Differential Phase Shift Keying (DPSK)

Differential phase shift keying (DPSK) is differentially coherent modulation method. DPSK does not need a synchronous (coherent) carrier at the demodulator. The input sequence of binary b its is modified such that the next bit depends upon the previous bit. Therefore in the receiver the previous received bits are used to detect the present bit.

2.5.1 DPSK Transmitter and Receiver

2.5.1.1 Transmitter / Generator of DPSK Signal

Fig. 2.5.1 shows the scheme to generate DPSK signal.

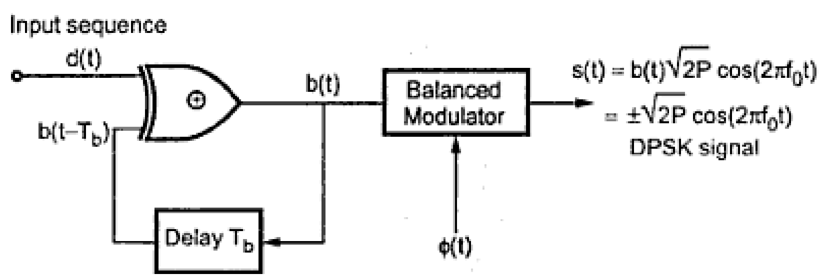


Fig. 2.5.1 Block diagram of DPSK generate or transmitter

The input sequence is $d(t)$. Output sequence is $b(t)$ and $b(t-T_b)$ is the previous output delayed by one bit period. Depending upon values of $d(t)$ and $b(t-T_b)$, exclusive OR gate generates the output sequence $b(t)$. Table 2.5.1 shows the truth table of this operation.

$d(t)$	$b(t-T_b)$	$b(t)$
0(−1 V)	0(−1 V)	0(−1 V)
0(−1 V)	1(1 V)	1(1 V)
1(1 V)	0(−1 V)	1(1 V)
1(1 V)	1(1 V)	0(−1 V)

Table 2.5.1 Truth table of exclusive OR gate

An arbitrary sequence $d(t)$ is taken. Depending on this sequence, $b(t)$ and $b(t-T_b)$ are found. These waveforms are shown in Fig. 2.5.2. The above Table 2.5.1 is used to derive the levels of these waveforms.

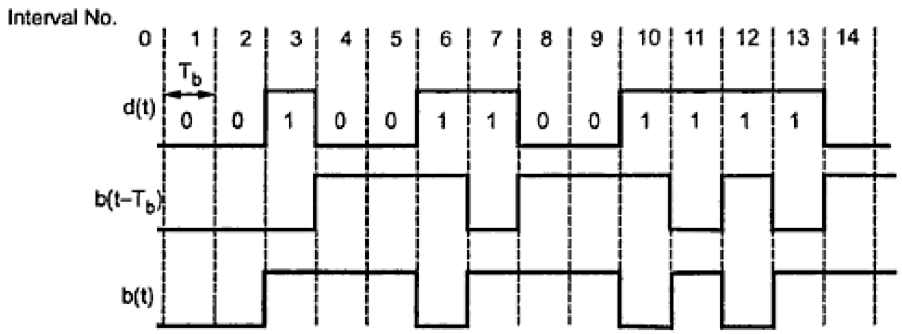


Fig. 2.5.2 DPSK waveforms

From the above waveform it is clear that $b(t-T_b)$ is the delayed version of $b(t)$ by one bit period T_b. The exclusive OR operation is satisfied in any interval i.e. in any interval $b(t)$ is given as,

$$b(t) = d(t) \oplus b(t-T_b) \qquad \text{... (2.5.1)}$$

While drawing the waveforms the value of $b(t-T_b)$ is not known initially in interval no.1. Therefore it is assumed to be zero and then waveforms are drawn. We can write some important conclusions from the waveforms

1. Output sequence $b(t)$ changes level at the beginning of each interval in which $d(t)=1$ and it does not changes level when $d(t)=0$. Observe that $d(3)=1$, hence level of $b(3)$ is changed at the beginning of interval 3. Similarly in intervals 10, 11, 12 and 13 $d(t)=1$. Hence $b(t)$ is changed at the starting of these intervals. In interval 8 and 9 $d(t)=0$. Hence $b(t)$ is not changed in these intervals.

2. When $d(t)=0$, $\quad b(t)=b(t-T_b)$ and
   When $d(t)=1$, $\quad b(t)=\overline{b(t-T_b)}$

3. In interval no.1. we have assumed $b(t-T_b)=0$ and we obtained the waveform as shown in Fig. 2.5.2. If we assume $b(t-T_b)=1$ in interval no. 1, then the waveform of $b(t)$ will be inverted. But still $b(t)$ changes the level at the beginning of each interval in which $d(t)=1$.

4. The sequence $b(t)$ modulates sinusoidal carrier.

5. When $b(t)$ changes the level, phase of the carrier is changed. Since $b(t)$ changes its level only if $d(t)=1$; It shows that phase of the carrier is changed only if $d(t)=1$. In PSK phase of the carrier changes on both the symbol '1' and '0'. Whereas in DPSK phase of the carrier changes only on symbol '1'. This is the main difference between PSK and DPSK.

6. Always two successive bits of $d(t)$ are checked for any change of level. Hence one symbol has two bits.

$\therefore$ Symbol duration (T) = Duration of two bits $(2T_b)$

i.e.
$$T = 2T_b \qquad \text{... (2.5.2)}$$

As shown in Fig. 2.5.1, the sequence $b(t)$ is applied to a balanced modulator. The balanced modulator is also supplied with a carrier $\sqrt{2P}\cos(2\pi f_0 t)$

The modulator output is,

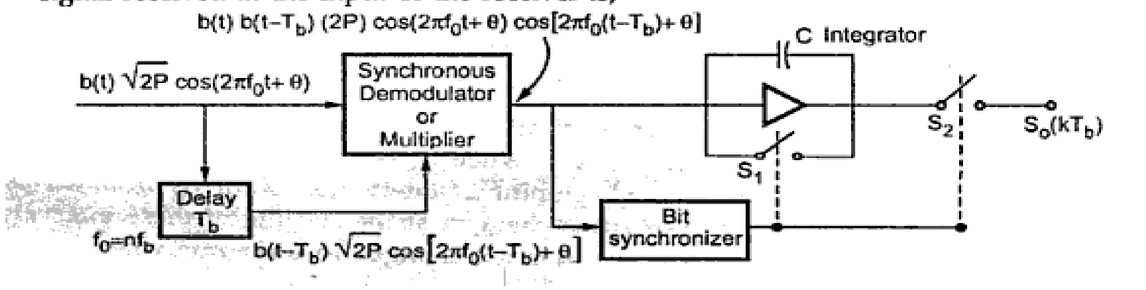
$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t) \qquad \text{... (2.5.3)}$$

$$= \pm\sqrt{2P}\cos(2\pi f_0 t) \qquad \text{... (2.5.4)}$$

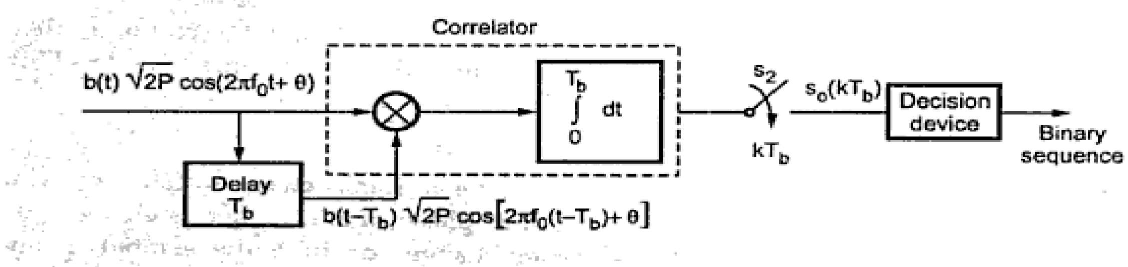
The above equation gives DPSK signal. Fig. 2.5.2 shows this DPSK waveforms. As shown in the waveforms the phase changes only when $d(t) = 1$.

2.5.1.2 DPSK Receiver

Fig. 2.5.3 shows the method to recover the binary sequence from DPSK signal. Fig. 2.5.3 (a) and (b) are equivalent to each other. Fig. 2.5.3(b) represents DPSK receiver using correlator. Fig. 2.5.3(a) shows multiplier and integrators separately. During the transmission, the DPSK signal undergoes some phase shift θ. Therefore the signal received at the input of the receiver is,



(a)



(b)

$$\text{Received signal} = b(t)\sqrt{2P}\cos(2\pi f_0 t+\theta) \qquad \text{... (2.5.5)}$$

This signal is multiplied with its delayed version by one bit. Therefore the output of the multiplier is,

$$\text{Multiplier output} = b(t)\,b(t-T_b)\,(2P)\cos(2\pi f_0 t+\theta)\cos\left[2\pi f_0 (t-T_b)+\theta\right] \qquad \text{... (2.5.6)}$$

We know that, $\cos(A)\cos(B) = \frac{1}{2}\left[\cos(A-B)+\cos(A+B)\right]$

Here $\quad A = 2\pi f_0 t+\theta \quad$ and $\quad B = 2\pi f_0 (t-T_b)+\theta$

$$\therefore \text{ Multiplier output} = b(t)\,b(t-T_b)\,P\left\{\cos 2\pi f_0 T_b + \cos\left[4\pi f_0\left(t-\frac{T_b}{2}\right)+2\theta\right]\right\} \qquad \text{...(2.5.7)}$$

f_0 is the carrier frequency and T_b is one bit period. T_b contains integral number of cycles of f_0. We know that,

$$f_b = \frac{1}{T_b}$$

If T_b contains 'n' cycles of f_0 then we can write,

$$f_0 = n f_b \quad \Rightarrow \quad f_0 = \frac{n}{T_b}$$

$$\therefore \qquad f_0 T_b = n \qquad \text{... (2.5.8)}$$

Putting $f_0 T_b = n$ in first cosine term in equation (2.5.7) we get

$$\text{Multiplier output} = b(t)\,b(t-T_b)\,P\left\{\cos 2\pi n + \cos\left[4\pi f_0\left(t-\frac{T_b}{2}\right)+2\theta\right]\right\}$$

This signal is given to the integrator. In the k^{th} bit interval, the integrator output can be written as,

$$s_o(kT_b) = b(kT_b)\, b\left[(k-1)T_b\right] P \int_{(k-1)T_b}^{kT_b} dt + b(kT_b)\, b\left[(k-1)T_b\right] P \int_{(k-1)T_b}^{kT_b} \cos\left[4\pi f_0\left(t - \frac{T_b}{2}\right) + 2\theta\right] dt$$

The integration of the second term will be zero since it is integration of carrier over one bit duration. The carrier has integral number of cycles over one bit period hence integration is zero. Therefore we can write,

$$\begin{aligned} s_o(kT_b) &= b(kT_b)\, b\left[(k-1)T_b\right] P\left[kT_b - (k-1)T_b\right] \\ &= b(kT_b)\, b\left[(k-1)T_b\right] P T_b \qquad \text{... (2.5.10)} \end{aligned}$$

Here know that $PT_b = E_b$; i.e. energy of one bit. The product $b(kT_b)\, b\left[(k-1)T_b\right]$ decides the sign of PT_b.

The transmitted data bit $d(t)$ can be verified easily from product $b(kT_b)\, b\left[(k-1)T_b\right]$. We know from Fig. 2.5.2 when $b(t) = b(t - T_b)$, $d(t) = 0$. That is if both are $+1\,V$ or $-1\,V$ then $b(t)\, b(t - T_b) = 1$. Alternately we can write,

If $\quad b(t)\, b(t - T_b) = 1\,V \quad$ then $\quad d(t) = 0$

We know that $b(t) = \overline{b(t - T_b)}$ then $d(t) = 1$. That is $b(t) = -1\ V, b(t - T_b) = +1\ V$ and vice versa. Therefore $b(t)\, b(t - T_b) = -1$. Alternately we can write,

If $\quad b(t)\, b(t - T_b) = -1\,V, \quad$ then $\quad d(t) = 1$

The decision device is shown in Fig. 2.5.3 (b). We know that,

$$s_o(kT_b) = b(kT_b)\, b\left[(k-1)T_b\right] P T_b \qquad \text{... from equation 2.5.10}$$

If $$s_o(kT_b) = \begin{cases} -PT_b, \text{ then } d(t) = 1 \text{ and} \\ +PT_b, \text{ then } d(t) = 0 \end{cases}$$

UNIT III DIGITAL TRANSMISSION

INTRODUCTION:

3.1 Pulse Modulation

The continuous time signal $x(t)$ to be transmitted is sampled at frequency f_s sufficiently above the highest frequen cy present in $x(t)$. The amplitude of the modulating signal $x(t)$ modulates some parameter of the pulse train. These parameters are amplitude, duration (width) and position. Fig. 3.1.1 shows different types of analog pulse modulation techniques with message waveform $x(t)$.

For PAM the modulated pulse parameter is amplitude, for PDM it is width and for PPM it is relative position. These parameters vary in direct proportion to amplitude of $x(t)$ at the sampling instant. As shown in waveforms of Fig. 3.1.1.

$$f_s = \frac{1}{T_s} = \text{Sampling frequency}$$

$$A_0 = \text{Amplitude of the pulse}$$

and $$\tau_0 = \text{Width of the pulse}$$

Since the waveforms are unipolar, they have some dc value. Also the shape of the pulse should be preserved (rising and falling edges, amplitude, duration etc.). Thus the transmission bandwidth needed for these pulse transmission is quite high compared to the message signal bandwidth. Therefore normally single channel PAM, PPM or PDM are seldom used. Always time division multiplexing (TDM) is used to utilize the channel bandwidth.

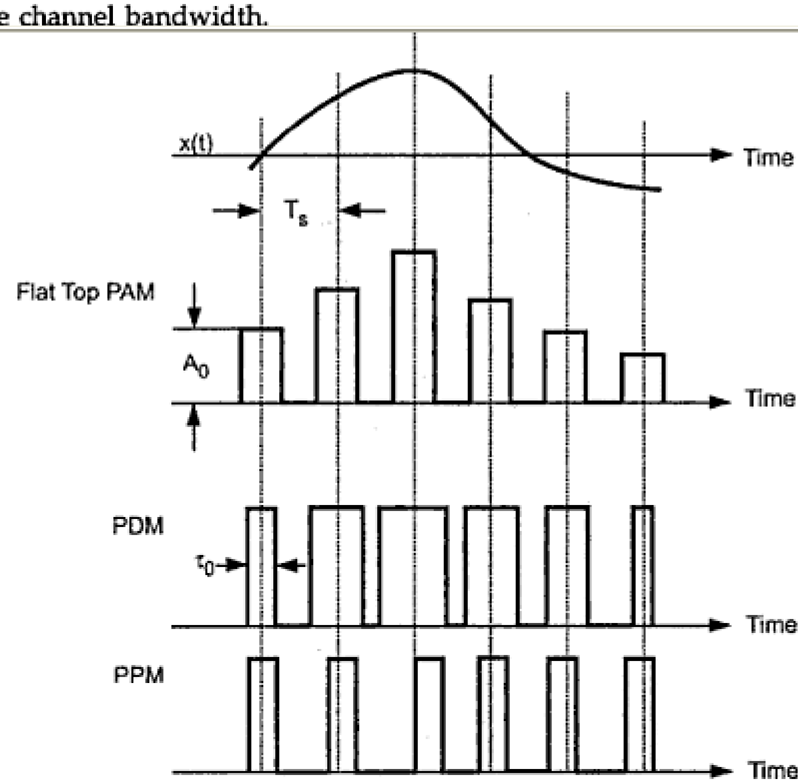


Fig. 3.1.1 Different types of analog pulse modulation techniques

3.2 Pulse Code Modulation (PCM)

3.2.1 PCM Generator

Nov./Dec.-2005 ; May/June - 2006

The pulse code modulator technique samples the input signal $x(t)$ at frequency $f_s \geq 2W$. This sampled 'Variable amplitude' pulse is then digitized by the analog to digital converter. The parallel bits obtained are converted to a serial bit stream. Fig. 3.2.1 shows the PCM generator.

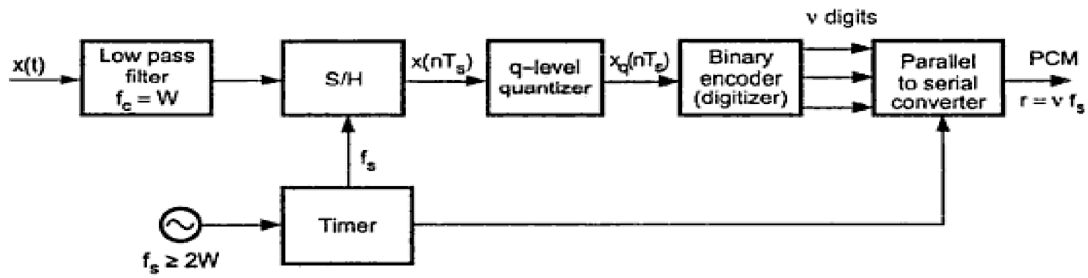


Fig. 3.2.1 PCM generator

In the PCM generator of above figure, the signal $x(t)$ is first passed through the low-pass filter of cutoff frequency 'W' Hz. This low-pass filter blocks all the frequency components above 'W' Hz. Thus $x(t)$ is bandlimited to 'W' Hz. The sample and hold circuit then samples this signal at the rate of f_s. Sampling frequency f_s is selected sufficiently above Nyquist rate to avoid aliasing i.e.,

$$f_s \geq 2W$$

In Fig. 3.2.1 output of sample and hold is called $x(nT_S)$. This $x(nT_S)$ is discrete in time and continuous in amplitude. A q-level quantizer compares input $x(nT_S)$ with its fixed digital levels. It then assigns any one of the digital level to $x(nT_S)$ which results in minimum distortion or error. This error is called *quantization error*. Thus output of quantizer is a digital level called $x_q(nT_S)$.

Quantization error is given as,

$$\varepsilon = x_q(nT_S) - x(nT_S) \qquad \text{... (3.2.1)}$$

3.2.2 Transmission Bandwidth in PCM

Let the quantizer use 'v' number of binary digits to represent each level. Then the number of levels that can be represented by 'v' digits will be,

$$q = 2^v \qquad \text{... (3.2.2)}$$

Here 'q' represents total number of digital levels of q-level quantizer.

For example if $v = 3$ bits, then total number of levels will be,

$$q = 2^3 = 8 \text{ levels}$$

Each sample is converted to 'v' binary bits. i.e. Number of bits per sample = v

We know that, Number of samples per second = f_s

∴ Number of bits per second is given by,

(Number of bits per second) = (Number of bits per samples) × (Number of samples per second)

= v bits per sample × f_s samples per second

3.2.3 PCM Receiver

Nov./Dec.-2005 ; May/June - 2006

Fig. 3.2.2 (a) shows the block diagram of PCM receiver and Fig. 3.2.2 (b) shows the reconstructed signal. The regenerator at the start of PCM receiver reshapes the pulses and removes the noise. This signal is then converted to parallel digital words for each sample.

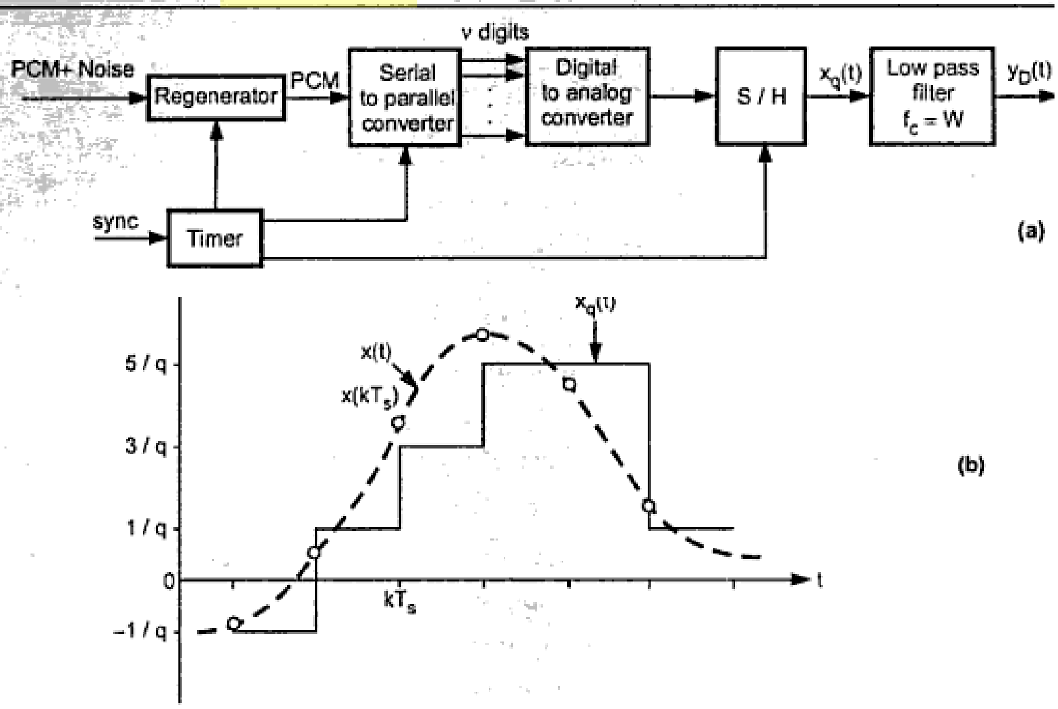


Fig. 3.2.2 (a) PCM receiver
(b) Reconstructed waveform

The digital word is converted to its analog value $x_q(t)$ along with sample and hold. This signal, at the output of S/H is passed through lowpass reconstruction filter to get $y_D(t)$. As shown in reconstructed signal of Fig. 3.2.2 (b), it is impossible to reconstruct exact original signal $x(t)$ because of permanent quantization error introduced during quantization at the transmitter. This quantization error can be reduced by increasing the binary levels. This is equivalent to increasing binary digits (bits) per sample. But increasing bits 'v' increases the signaling rate as well as transmission bandwidth as we have seen in equation (3.2.4) and equation (3.2.7).

1.8.6.3 Companding in PCM

Normally we don't know how the signal level will vary in advance. Therefore the nonuniform quantization (variable step size 'δ') becomes difficult to implement. Therefore the signal is amplified at low signal levels and attenuated at high signal levels. After this process, uniform quantization is used. This is equivalent to more step size at low signal levels and small step size at high signal levels. At the receiver a reverse process is done. That is signal is attenuated at low signal levels and amplified at high signal levels to get original signal. Thus the compression of signal at transmitter and expansion at receiver is called combinely as *companding*. Fig. 1.8.9 shows compression and expansion curves.

As can be seen from Fig. 1.8.9, at the receiver, the signal is expanded exactly opposite to compression curve at transmitter to get original signal. A dotted line in the Fig. 1.8.9 shows uniform quantization. The compression and expansion is obtained by passing the signal through the amplifier having nonlinear transfer characteristic as

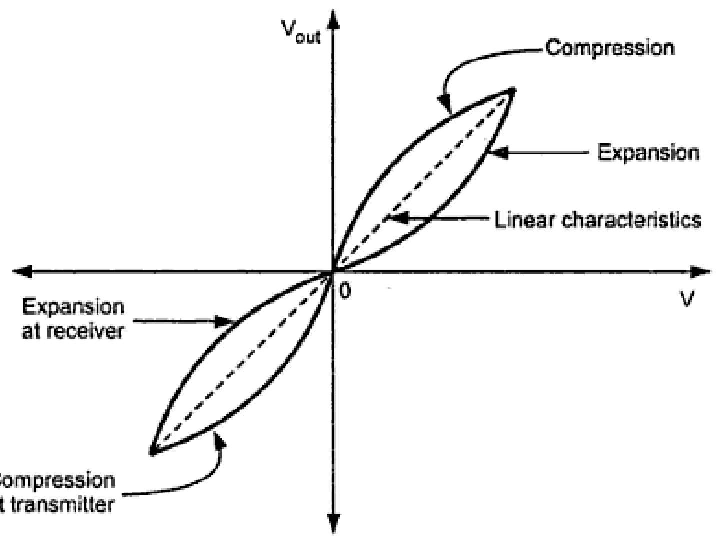


Fig. 1.8.9 Companding curves for PCM

1.8.6.4 μ - Law Companding for Speech Signals

Normally for speech and music signals a μ - law compression is used. This compression is defined by the following equation,

$$Z(x) = (\text{Sgn } x) \frac{ln(1+\mu|x|)}{ln(1+\mu)} \quad |x| \le 1 \qquad \dots (1.8.52)$$

Fig. 1.8.10 shows the variation of signal to noise ratio with respect to signal level without companding and with companding.

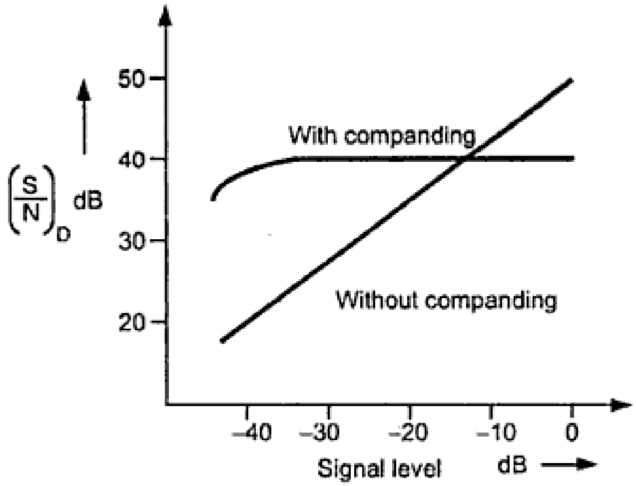


Fig. 1.8.10 PCM performance with μ - law companding

1.8.6.5 A-Law for Companding

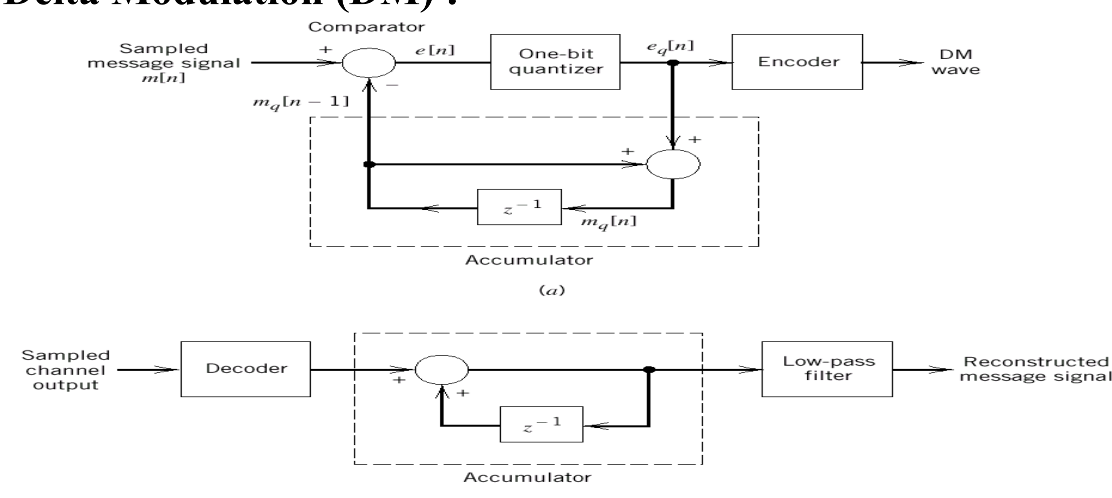
The A law provides piecewise compressor characteristic. It has linear segment for low level inputs and logarithmic segment for high level inputs. It is defined as,

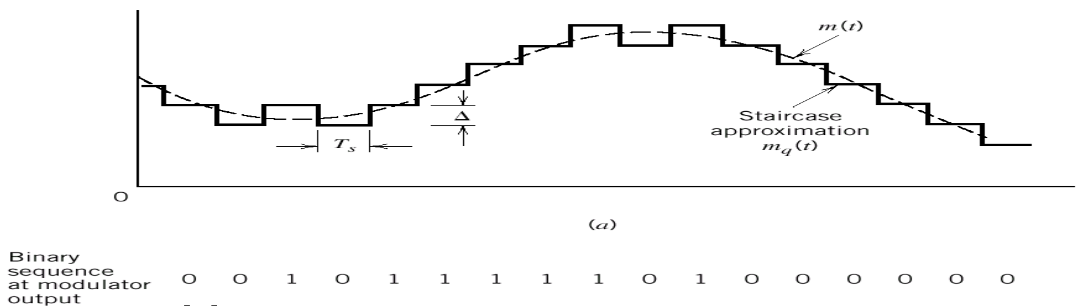
$$Z(x) = \begin{cases} \dfrac{A|x|}{1+ln A} & for \ 0 \leq |x| \leq \dfrac{1}{A} \\ \dfrac{1+ln(A|x|)}{1+ln A} & for \ \dfrac{1}{A} \leq |x| \leq 1 \end{cases} \quad \text{... (1.8.53)}$$

When A = 1, we get uniform quantization. The practical value for A is 87.56. Both A-law and μ-law companding is used for PCM telephone systems.

Delta Modulation (DM) :

Delta Modulation (DM) :





Let $m[n] = m(nT_s)$, $n = 0, \pm 1, \pm 2, \ldots$ (b)

where T_s is the sampling period and $m(nT_s)$ is a sample of $m(t)$.

The error signal is

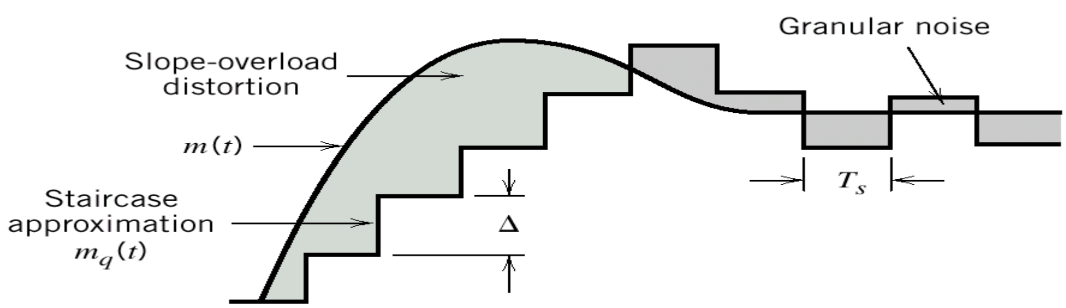
$e[n] = m[n] - m_q[n-1]$ (3.52)

$e_q[n] = \Delta \operatorname{sgn}(e[n])$ (3.53)

$m_q[n] = m_q[n-1] + e_q[n]$ (3.54)

where $m_q[n]$ is the quantizer output , $e_q[n]$ is the quantized version of $e[n]$, and Δ is the step size

The modulator consists of a comparator, a quantizer, and an accumulator. The output of the accumulator is



Differential Pulse-Code Modulation (DPCM):

Usually **PCM** has the sampling rate higher than the **Nyquist rate** .The encode signal contains redundant information. **DPCM** can efficiently remove this redundancy**.**

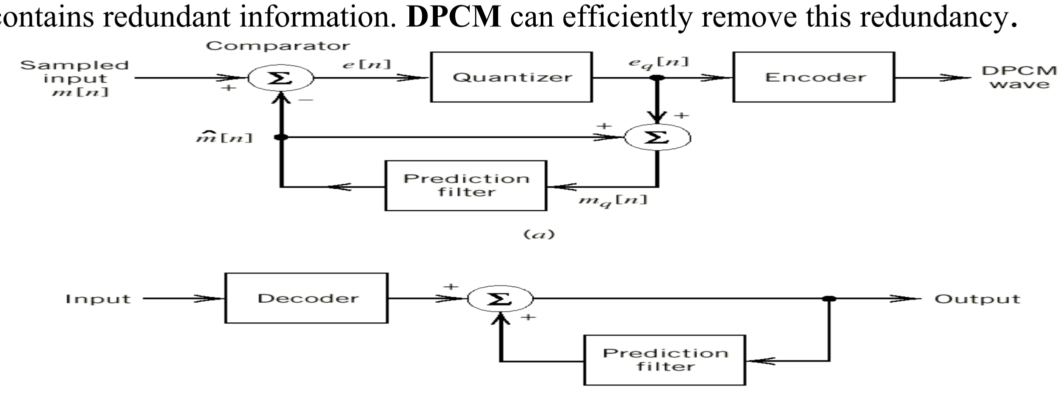


Figure 3.28 DPCM system. (*a*) Transmitter. (*b*) Receiver.

$$e[n]=m[n]-\hat{m}[n] \qquad (3.74)$$

$\hat{m}[n]$ is a prediction value.

The quantizer output is

$$e_q[n]=e[n]+q[n] \qquad (3.75)$$

where $q[n]$ is quantization error.

The prediction filter input is

$$m_q[n]=\underbrace{\hat{m}[n]+e[n]}_{m[n]}+q[n] \qquad (3.77)$$

$$\Rightarrow m_q[n]=m[n]+q[n] \qquad (3.78)$$

Processing Gain:

The $(\text{SNR})_o$ of the DPCM system is

$$(\text{SNR})_o=\frac{\sigma_M^2}{\sigma_Q^2} \qquad (3.79)$$

where σ_M^2 and σ_Q^2 are variances of $m[n]\,(E[m[n]]=0)$ and $q[n]$

$$(\text{SNR})_o=\left(\frac{\sigma_M^2}{\sigma_E^2}\right)\left(\frac{\sigma_E^2}{\sigma_Q^2}\right)$$

$$=G_p(\text{SNR})_Q \qquad (3.80)$$

where σ_E^2 is the variance of the predictions error

and the signal-to-quantization noise ratio is

$$(\text{SNR})_Q=\frac{\sigma_E^2}{\sigma_Q^2} \qquad (3.81)$$

Processing Gain, $G_p=\dfrac{\sigma_M^2}{\sigma_E^2}$ (3.82)

Design a prediction filter to maximize G_p (minimize σ_E^2)

NEAR EAST UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

COM 318

DATA COMMUNICATIONS

LECTURE NOTES

Prepared by

Dr. Tayseer Alshanableh

Nicosia-2007

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO DATA COMMUNICATIONS

- **COMPUTER NETWORK**

Interconnected collection of autonomous computers that are able to exchange information.

- No master/slave relationship between computers in the network.

- **DATA COMMUNICATIONS**

Transmission of signals in a reliable and efficient matter.

- **COMMUNICATION MODEL (SYSTEM)**

The purpose of a communications system is to exchange data between two entities.

- **Source:** entity that generates data; eg. a person who speaks into the phone, or a computer sending data to the modem.
- **Transmitter:** a device to transform/encode the signal generated by the source.
  - the transformed signal is actually sent over the transmission system.

  eg. a modem transforms digital data to analog signal that can be handled by the telephone network.
- **Transmission System (Channel):** medium that allows the transfer of a signal from one point to another.

  eg. a telephone network for a computer/modem.
- **Receiver:** a device to decode the received signal for handling by destination device.

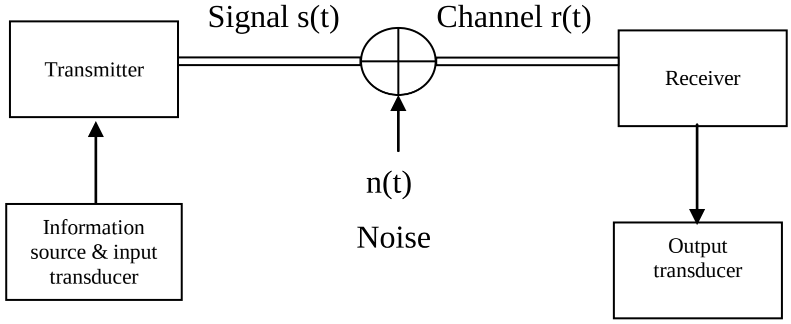
  eg. A modem converts the received analog data back to digital for the use by the computer.
- **Destination:** entity that finally uses the data.

  eg. Computer on other end of a receiving modem.

Data Communications

Data communications is the transfer of information that is in digital form, before it enters the communication system.

- Basic Elements of a Communication System



Basic Elements of a Communication System

- **Information:** generated by the source may be in the form of voice, a picture or a plain text. An essential feature of any source that generates information is that its output is described in probabilistic terms; that is, the output is not deterministic.
  A transducer is usually required to convert the output of a source in an electrical signal that is suitable for transmission.

- **Transmitter:** a transmitter converts the electrical signal into a form that is suitable for transmission through the physical channel or transmission medium. In general, a transmitter performs the matching of the message signal to the channel by a process called modulation.
  The choice of the type of modulation is based on several factors, such as:
  - the amount of bandwidth allocated,
  - the type of noise and interference that the signal encounters in transmission over the channel,
  - and the electronic devices that are available for signal amplification prior to transmission.

- **Channel:** the communication channel is the physical medium that connects the transmitter to the receiver. The physical channel may be a pair of wires that carry the electrical signals, or an optical fibre that carries the information on a modulated light beam or free space at which the information-bearing signal are electromagnetic waves.

- **Receiver:** the function of a receiver is to recover the message signal contained in the received signal. The main operations performed by a receiver are demodulation, filtering and decoding.

Analog and Digital Data Transmission

- Data are entries that convey information.
- Signals are electrical encoding (representation) of data.
- Signalling is the act of propagation of signals through a suitable medium.

The terms analog and digital correspond to continuous and discrete, respectively. These two terms are frequently used in data communications.

Analog data takes on continuous values on some interval. The most familiar example of analog data is audio signal. Frequency components of speech may be found between

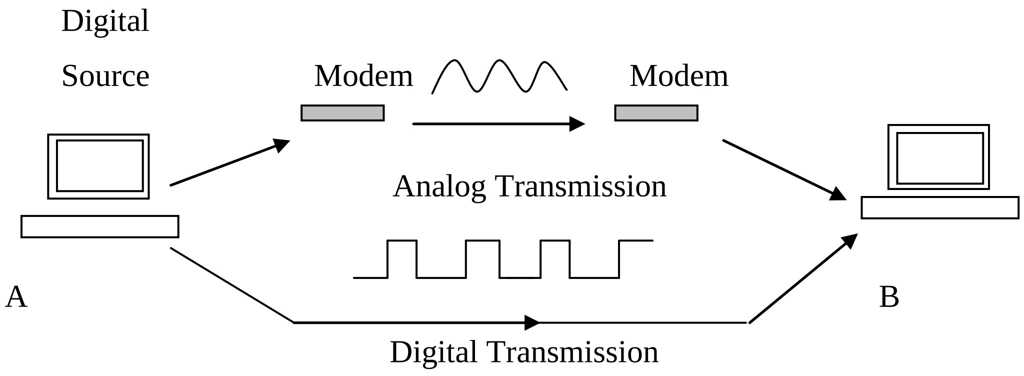
20 Hz and 20 kHz. The basic speech energy is concentrated between 300-3400 Hz. The frequencies up to 4000 Hz add very little to the intelligibility of human ear.

Another common example of analog data is video. The outputs of many sensors, such as temperature and pressure sensors, are also examples of analog data.

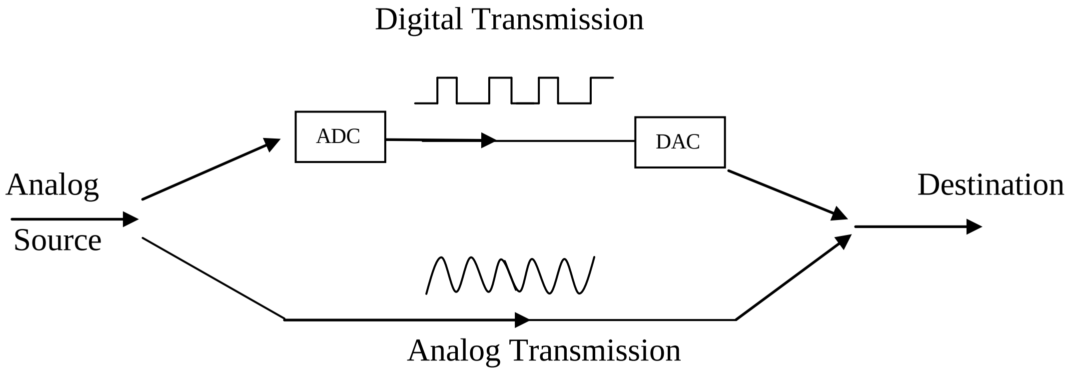
Digital data takes on discrete values; eg. a computer's output.

- Analog transmission is a means of transmitting analog signals regardless of their content. The data may be analog or digital.
- Digital transmission is the transfer of information through a medium in digital form. A digital signal can be transmitted only for a limited distance.
- Data communications is the transfer of information that is in digital form, before it enters the communication system.

- Two methods of sending data from computer A to computer B. both cases are examples of data communications, because the original data is digital in nature.



- Two ways of transmitting analog information. In either cases it is not data communications, because the original information is not digital.



ADC: Analog-Digital-Converter

DAC: Digital-Analog-Converter

Digital Communication System

Up to this point, we have described an electrical communication system in rather broad terms based on the implicit assumption that the message signal is a continuous time-varying waveform. We refer to such continuous-time signal waveforms as analog signals and to the corresponding information sources that produce such signals as analog sources. Analog signals can be transmitted directly via carrier modulation over the communication channel and demodulated accordingly at the receiver. We call such a communication system an analog communication system.

Alternatively, an analog source output may be converted into a digital form and the message can be transmitted via digital modulation and demodulated as a digital signal at the receiver. There are some potential advantages to transmitting an analog signal by means of digital modulation. The most important reason is that signal fidelity is better controlled through digital transmission than analog transmission. In particular, digital transmission allows us to regenerate the digital signal in long-distance transmission, thus eliminating effects of noise at each regeneration point. In contrast, the noise added in analog transmission is amplified analog with the signal when amplifiers are used periodically to boost the signal level in long-distance transmission. Another reason for choosing digital transmission over analog is that the analog message signal may be highly redundant. With digital processing, redundancy may be removed prior to modulation, thus conserving channel bandwidth. Yet a third reason may be that digital communication systems are often cheaper to implement.

In some applications, the information to be transmitted is inherently digital, e.g., in the form of English text, computer data, etc. In such cases, the information source that generates the data is called a discrete (digital) source.

In a digital communication system, the functional operations performed at the transmitter and receiver must be expanded to include message signal discrimination at the transmitter and message signal synthesis or interpolation at the receiver. Additional functions include redundancy removal, and channel coding and decoding.

Figure 1.2 illustrates the functional diagram and the basic elements of a digital communication system. The source output may be either an analog signal, such as audio or video signal, or a digital signal, such as the output of a Teletype machine, which is discrete in time and has a finite number of output characters. In a digital communication system, the messages produced by the source are usually converted into a sequence of binary digits.

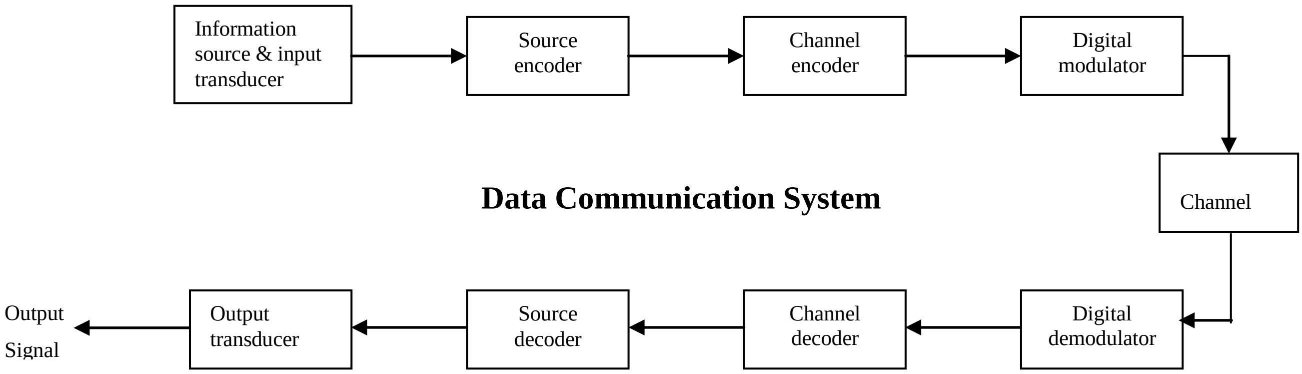


Figure 1.2. Basic elements of a digital communication system

Ideally, we would like to represent the source output (message) by as few binary digits as possible. In other words, we seek an efficient representation of the source output that results in little or no redundancy. The process of efficiently converting the output of either an analog or a digital source into a sequence of binary digits is called source encoder or data compression.

The sequence of binary digits from the source encoder, which we call the information sequence, is passed to the channel encoder. The purpose of the channel encoder is to introduce in a controlled manner some redundancy in the binary information sequence, which can be used at the receiver to overcome the effects of noise and interference encountered in the transmission of the signal through the channel. Thus, the added redundancy serves to increase the reliability of the received data and improves the fidelity of the received signal. In effect, redundancy in the information sequence aids the receiver in decoding the desired information sequence. For example, a (trivial) form of encoding of the binary information sequence is simply to repeat each binary digit m times, where m is some positive integer. More sophisticated (nontrivial) encoding involves taking k information bits at a time and mapping each k-bit sequence into a unique n-bit sequence, called a code word. The amount of redundancy introduced by encoding the data in this manner is measured by the ratio n/k. The reciprocal of this ratio, namely, k/n is called the rate of the code or, simply, the code rate.

The binary sequence at the output of the channel encoder is passed to the digital modulator, which serves as the interface to the communications channel. Since nearly all of the communication channels encountered in practice are capable of transmitting electrical signals (waveforms), the primary purpose of the digital modulator is to map the binary information sequence into signal waveforms. To elaborate on this point, let us suppose that the coded information sequence is to be transmitted one bit at a time at some uniform rate R bits/s. The digital modulator may simply map the binary digit 0 into a waveform $s_0(t)$ and the binary digit 1 into a waveform $s_1(t)$. In this manner, each bit from the channel encoder is transmitted separately. We call this binary modulation. Alternatively, the modulator may transmit b coded information bits at a time by using M =2b distinct waveform $s_i(t)$, I = 0, 1, …, m-1, one waveform for each of the 2b possible b-bits sequences. We call this M-ary modulation (M >2). Note that a new b-bit sequence enters the modulator every b/R seconds. Hence, when the channel bit rate R is fixed, the amount of time available to transmit one of the M waveforms corresponding to a b-bit sequence is b times the period in a system that uses binary modulation.

At the receiving end of a digital communications system, the digital demodulator processes the channel-corrupted transmitted waveform and reduces each waveform to a single number that represents an estimate of the transmitted data symbol. For example, when binary modulation is used, the demodulator may process the received waveform and decide on whether the transmitted bit is a 0 or 1. In such a case, we say the demodulator has made a binary decision. As one alternative, the demodulator may make a ternary decision; that is, it decides that the transmitted bit is either a 0 or 1 or it makes no decision at all, depending on the apparent quality of the received signal. When no decision is made on a particular bit, we say that the demodulator has inserted an erasure in the demodulated data. Using the redundancy in the transmitted data, the decoder attempts to fill in the positions where erasures occurred. Viewing the decision process performed by the demodulator as a form of quantization, we observe that binary and ternary decisions are special cases of a demodulator that quantizes to Q levels, where $Q \geq 2$ In general, if the digital communications system employs M-ary modulation, where m = 0,1, … , M represent the M possible transmitted symbols, each corresponding to $k = \log_2 M$ bits, the demodulator may make A Q-ary decision, where $Q \geq M$ .In the extreme case where no quantization is performed, $Q = \infty$.

When there is no redundancy in the transmitted information, the demodulator must decide which of the M waveforms was transmitted in any given time interval. Consequently, Q = M, and since there is no redundancy in the transmitted information, no discrete channel decoder is used following the demodulator. On the other hand, when there is redundancy introduced by a discrete channel encoder at the transmitter, the Q-ary output from the demodulator occurring every k/R seconds is fed to the decoder, which attempts to reconstruct the original information sequence from knowledge of the code used by the channel encoder and the redundancy contained in the received data. A measure of how well the demodulator and encoder perform is the frequency with which errors occur in the decoded sequence. More precisely, the average probability of a bit-error at the output of the decoder is a measure of the performance of the demodulator-decoder combination. In general, the probability of error is a function of the code characteristics, the types of waveforms used to transmit the information over the channel, the transmitter power, the characteristics of the channel (i.e., the amount of noise), the nature of the interference, etc., and the method of demodulation and decoding. These items and their effect on performance will be discussed in detail in subsequent chapters.

As a final step, when an analog output is desired, the source decoder accepts the output sequence from the channel decoder, and from knowledge of the source encoding method used, attempts to reconstruct the original signal from the source. Due to channel decoding errors and possible distortion introduced by the source encoder and, perhaps, the source decoder, the signal at the output of the source decoder is an approximation to the original source output. The difference or some function of the difference between the original signal and the reconstructed signal is a measure of the distortion introduced by the digital communications system.

Early Work in Digital Communications

Although Morse is responsible for the development of the first electrical digital communication system (telegraphy), the beginnings of what we now regard as modem digital communications stem from the work of Nyquist (1924), who investigated the problem of determining the maximum signalling rate that can be used over a telegraph channel of a given bandwidth without intersymbol interference. He formulated a model of a telegraph system in which a transmitted signal has the general form

$$S(t) = \sum_n a_n g(t - nT)$$

Where $g(t)$ represents a basic pulse shape and $\{a_n\}$ is the binary data sequence of $\{\pm 1\}$ transmitted at a rate of $1/T$ bits per second. Nyquist set out to determine the optimum pulse shape that was bandlimited to W Hz and maximised the bit rate $1/T$ under the constraint that the pulse caused no intersymbol interference at the sampling times k/T, $k = 0, \pm 1, \pm 2, \ldots$. His studies led him to conclude that the maximum pulse rate$1/T$ is $2W$ pulses per second. This rate is now called the Nyquist rate. Moreover, this pulse rate can be achieved by using the pulses $g(t) = (\sin 2\pi\, Wt)/2\pi\, Wt$. This pulse shape allows the recovery of the data without intersymbol interference at the sampling instants Nyquist's result is equivalent to a version of the sampling theorem for bandlimited signals, which was later stated precisely by Shannon (1948). The sampling theorem states that a signal of bandwidth W can be reconstructed from samples taken at the Nyquist rate of $2W$ samples per second using the interpolation formula

$$S(t) = \sum_n s\left(\frac{n}{2W}\right) \frac{\sin 2\pi W(t - n/2W)}{2\pi W(t - n/2W)}$$

In light of Nyquist's work Hartley (1928) considered the issue of the amount of data that can be transmitted reliably over a bandlimited channel when multiple amplitude levels are used. Due to the presence of noise and other interference, Hartley postulated that the receiver could reliably estimate the received signal amplitude to some accuracy, say A_δ. This investigation led Hartley to conclude that there is maximum data rate that can be communicated reliably over a bandlimited channel when the maximum signal amplitude is limited to A_{max} (fixed power constraint) and the amplitude resolution is A_δ.

Another significant advance in the development of communications was the work of Wiener (1942) who considered the problem of estimating a desired signal waveform $s(t)$ in the presence of additive noise $n(t)$, based on observation of the received signal $r(t) = s(t) + n(t)$. This problem arises in signal demodulation. Wiener determined the linear filter whose output is the best mean-square approximation to the desired signal $s(t)$. The resulting filter is called the optimum linear (Wiener) filter. Hartley's and Nyquist results on the maximum transmission rate of digital information were precursors to the work of Shannon (1948 a, b) who established the mathematical foundations for information theory and derived the fundamental limits for digital communication systems. In his pioneering work, Shannon formulated the basic problem of reliable transmission of information in statistical terms, using probabilistic models for information sources and communication channels. Based on such a statistical formulation, he adopted a logarithmic measure for the information content of a source. He also demonstrated that the effect of a transmitter power constraint, a bandwidth constraint, and additive noise can be associated with the channel and incorporated into a single parameter, called the channel capacity For example, in the case of an additive white (spectrally flat) Gaussian noise interference, an ideal bandlimited channel of bandwidth W has a capacity C given by

$$C = W \log_2\left(1 + \frac{P}{WN_0}\right) \text{ bits/s}$$

where P is the average transmitted power and N_0 is the power spectral density of the additive noise. The significance of the channel capacity is as follows: If the information rate R from the source is less than C ($R < C$), then it is theoretically possible to achieve reliable (error-free) transmission through the channel by appropriate coding. On the other hand, if $R > C$, reliable transmission is not possible regardless of the amount of signal processing performed at the transmitter and receiver. Thus, Shannon established basic limits on communication of information and gave birth to a new field that is now called information theory.

Initially the fundamental work of Shannon had a relatively small impact on the design and development of new digital communications systems. In part, this was due to the small demand for digital information transmission during the 1950's. Another reason was the relatively large complexity and, hence, the high cost of digital hardware required to achieve the high efficiency and high reliability predicted by Shannon's theory.

Another important contribution to the field of digital communications is the work of Kotelnikov (1947), which provided a coherent analysis of the various digital communication systems based on a geometrical approach. Kotelnikov approach was later expanded by Wozencraft and Jacobs (1965).

The increase in the demand for data transmission during the last three decades, coupled with the development of more sophisticated integrated circuits, has led to the development of very efficient and more reliable digital communications systems. In the course of these developments, Shannon's original results and the generalization of his results on maximum transmission limits over a channel and on bounds on the performance achieved have served as benchmarks for any given communications system design. The theoretical limits derived by Shannon and other researchers that contributed to the development of information theory serve as an ultimate goal in the continuing efforts to design and develop more efficient digital communications systems.

Following Shannon's publications name the classic work of Hamming (1950) on error detecting and error-correcting codes to combat the detrimental effects of channel noise. Hamming's work stimulated many researchers in the years that followed, and a variety of new and powerful codes were discovered, many of which are used today in the implementation of modem communication systems.

CHAPTER II

DATA TRANSMISSION & SIGNALS

Data Transmission

Concepts and Terminology

- **Transmission Terminology**

  Transmission from transmitter to receiver goes over some transmission medium using electromagnetic waves.

  - Guided Media: waves are guided along a physical path; twisted pair, optical fibre, coaxial cable.

  - Unguided Media: waves are not guided; air waves radio waves.

  - Direct Link: signal goes from transmitter to receiver without intermediate devices, other than amplifiers and repeaters.

  - Point-to Point Link: guided media with direct link between two devices.

  - Multipoint Guided Configuration: more than two devices can share the same medium.

- **Frequency, Spectrum, & Bandwidth**

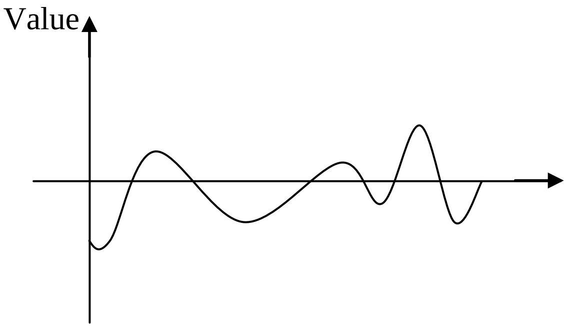
- Signal is generated by a transmitter and transmitted over a medium.
- Signal is a function of time or frequency.

A signal is any function that carries information. Based on the range of variation of independent variables, signals can be divided into two classes: continuous-time (or analog) signals and discrete-time (or digital) signals. A signal is a function of time, but can also be expressed as a function of frequency; that is, the signal consists of components of different frequencies.

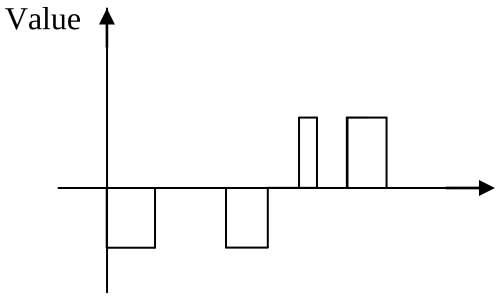
- **Analog and Digital Signals**

Information can be analog or digital. Analog information is continuous. Digital information is discrete.

Signals can be analog or digital. Analog signals can have any value in a range; while digital signals can have only a limited number of values.



Analog Signal



Digital Signal

- **Time –Domain Concepts**

- **Continuous Signal:** Signal intensity varies in a smooth fashion over time; no breaks or discontinuities in the signal.
- **Discrete Signals:** Signal intensity can take one of two pre-specified values for any amount of time.

A continuous time signal is defined by a continuous independent variable. A signal $s(t)$ is continuous if

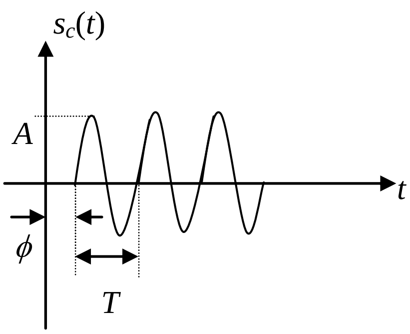
$$\lim_{t \to a} s_c(t) = s(t) \qquad \text{for all a.}$$

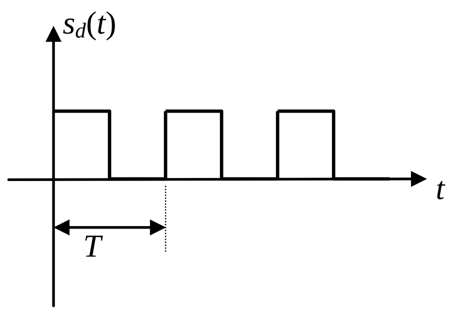
- **Periodic Signal**

A signal $s(t)$ is periodic if

$$s(t+T) = s(t)$$

where T is the period of the signal.





Three important characteristics of a periodic signal are: amplitude, frequency, and phase. Amplitude (A) is the instantaneous value of a signal at any time, and is measured in volts. Frequency (f) is the inverse of the period (T); ($f=1/T$), or the number of period repetition in one second, and is measured in cycles per second or Hertz (Hz). Phase (ϕ) is a measure of the relative position in time within a single period of a signal. Thus, we can express a sinusoid signal as

$$s(t) = A\sin(2\pi ft + \phi)$$

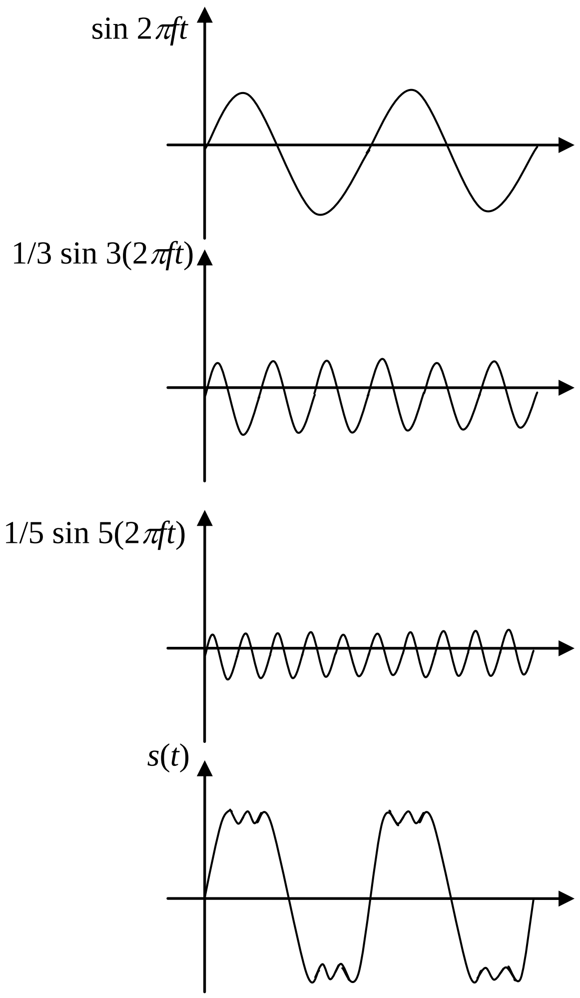
where A is the amplitude, f is the frequency, and ϕ is the phase.

- **Frequency-Domain Concepts**

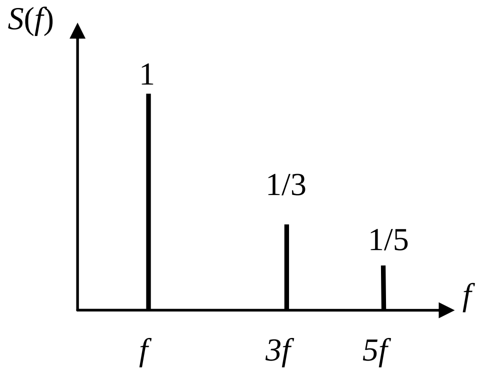
Any signal can also be viewed as a function of frequency, for example, the signal

$$s(t) = \sin 2\pi ft + 1/3\sin 3(2\pi ft) + 1/5\sin 5(2\pi ft)$$

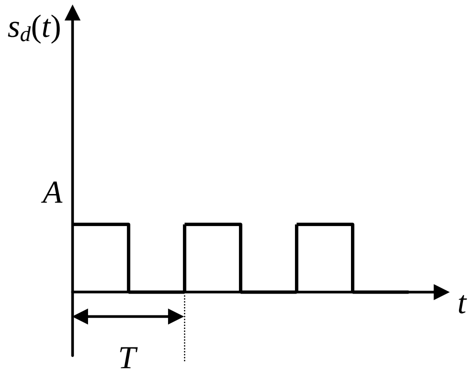
consists of three components as shown in the figure below:



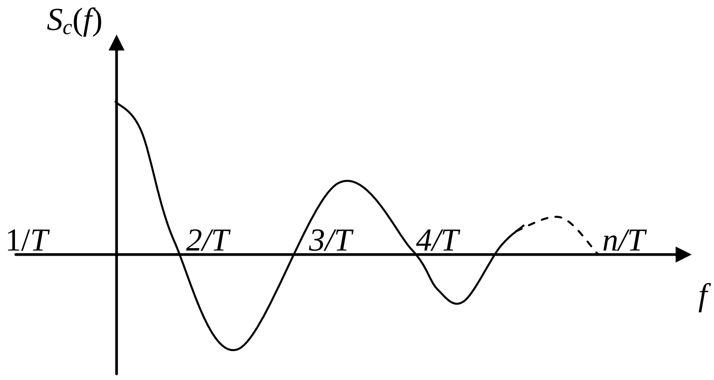
The frequency components of a signal can be determined using Fourier analysis. The following figure shows the spectrum $S(f)$ of the signal $s(t)$. The spectrum of a signal is the range of frequencies that it contains. For this signal the spectrum extends from f to $5f$. the spectrum in this case is discrete.



Many signals, such as the one in the following figure, have continuous spectrum $S_c(f)$ and

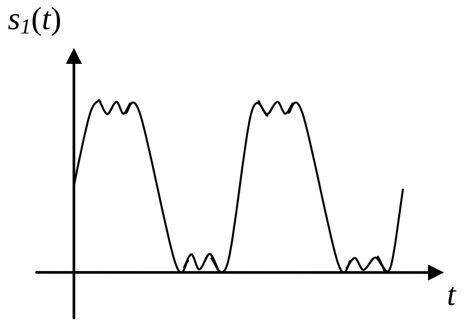


and an infinite bandwidth as shown below:

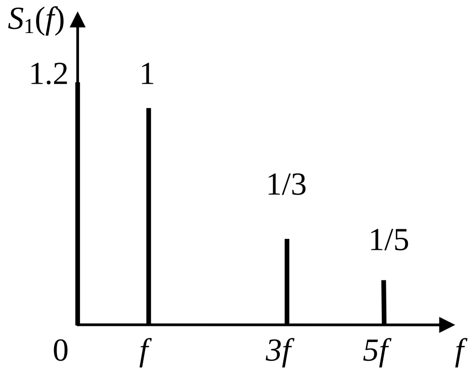


However, most of the energy in the signal is contained in a relatively narrow band of frequencies. This band is referred to as the effective bandwidth, or just bandwidth.
If a signal includes a component of zero frequency, that component is called d_c component or constant component.

The signal $s_1(t)$ in the following figure is obtained by adding a d_c component on $s(t)$:



With a d_c component, it has a frequency term at $f = 0$ and a non-zero average amplitude.



Spectrum of $s_1(t)$

$$s_1(t) = 1.2 + \sin(2\pi ft) + 1/3 \sin 3(2\pi ft) + 1/5 \sin 5(2\pi ft)$$

- **Fundamental Frequency**

Base frequency such that the frequency of all components can be expressed as its integer multiples; the period of the aggregate signal is the same as the period of the fundamental frequency:

- Each signal can be decomposed into a set of sinusoid signals by making use of Fourier's analysis.
- The time-domain function $s(t)$ specifies a signal in terms of its amplitude at each instant of time.
- The frequency-domain function $S(f)$ specifies the signal in terms of peak amplitude of constituent frequencies.

Spectrum

Range of frequencies contained in a signal.

Absolute Bandwidth

Width of the spectrum.

Effective Bandwidth

Narrow band of frequencies containing most of the energy of the signal.

DC Component

Component of zero frequency; changes the average amplitude of the signal to non-zero.

Relationship between Data Rate and Bandwidth

- Any transmitter/receiver system can accommodate only a limited range of frequencies.
  * The range for FM radio transmission is 88-108 MHz
- This limits the data rate that can be carried over the transmission medium.
- Consider a square wave. Suppose that we let the positive pulse to be binary 1 and the negative pulse to be binary 0. Then, the waveform represents the binary stream 1010… and duration (period) of each pulse is $1/2f$. Thus, the data rate is equal to $2f$ bits per second (bps) or the data rate is equal to twice the fundamental frequency of the digital signal. It can be shown that the frequency-domain representation of this waveform is:

$$s(t) = \sum_{k=1}^{\infty} \frac{1}{k} \sin(2\pi k f t)$$

- This waveform has infinite number of frequency components and infinite bandwidth.
- Peak amplitude of the k^{th} frequency component is $1/k$, so most of the energy is concentrated in the first few frequencies.

Ex

Consider a digital transmission system capable of transmitting signals with a bandwidth of 4 MHz.

Case 1

Approximate the square wave with a waveform of the first three sinusoidal components

$$\sin(2\pi f t) + 1/3\sin(2\pi(3f)t) + 1/5\sin(2\pi(5f)t)$$

If $f = 10^6$ cycles per second, or 1 MHz, the bandwidth of the signal

$$s(t) = \left[\sin(2\pi \times 10^6 t) + 1/3\sin(2\pi \times 3 \times 10^6 t) + 1/5\sin(2\pi \times 5 \times 10^6 t)\right]$$

is $5\text{x}10^6 - 10^6 = 4$ MHz

For f = 1 MHz, the period of the fundamental frequency is T = 1/10^6 = 1μs. If the waveform is a bit string of 1's and 0's, then one bit occurs every 0.5 μs for a data rate of 2x10^6 bps or 2 Mbps.

Case 2

Assume a bandwidth of 8 MHz and f = 2 MHz; this gives us the signal bandwidth as

$$(5\text{x}2\text{x}10^6)-(2\text{x}10^6) = 8 \text{ MHz}$$

But T = 1/f = 0.5 μs, so that the time needed for one bit is 0.25 μs, giving a data rate of 4 Mbps. Other things being equal, doubling the bandwidth doubles the potential data rate.

Case 3

Let us represent the signal by the first two components of the sinusoid as

$$\left[\sin(2\pi ft) + 1/3\sin(2\pi(3f)t)\right]$$

Assume that f = 2 MHz and T = 1/f = 0.5 μs so that the time needed for one bit is 0.25 μs, giving a data rate of 4 Mbps.

Bandwidth of the signal is

$$(3\text{x}2\text{x}10^6)-(2\text{x}10^6) = 4 \text{ MHz}$$

A given bandwidth can support various data rates depending on the ability of the receiver to differentiate between 0 and 1 in the presence of noise and other impairments.

Ex

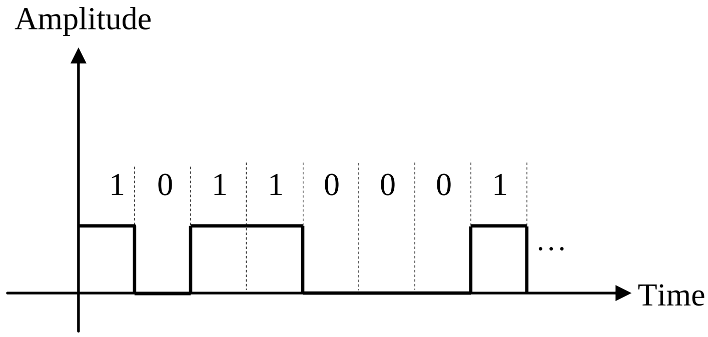
If a periodic signal is decomposed into five waves with frequencies of 100, 300, 500, 700, and 900 Hz, what is the bandwidth of the signal?

Let f_h be the highest frequency, f_l be the lowest frequency, and B be the bandwidth, then

$$B = f_h - f_l$$
$$= 900-100 = 800 \text{ Hz}$$

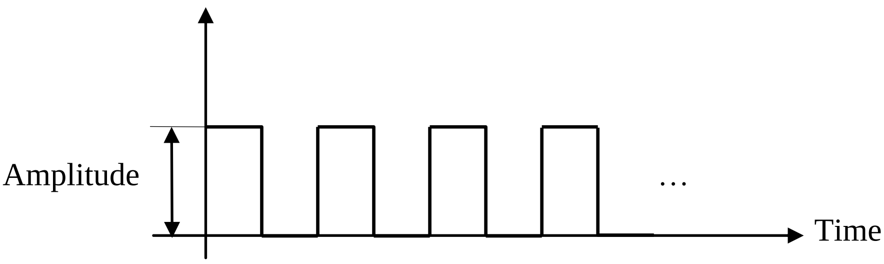
Digital Signals

Data can be represented by a digital signal. For example, a 1 can be encoded as a positive voltage and a 0 as a zero voltage.

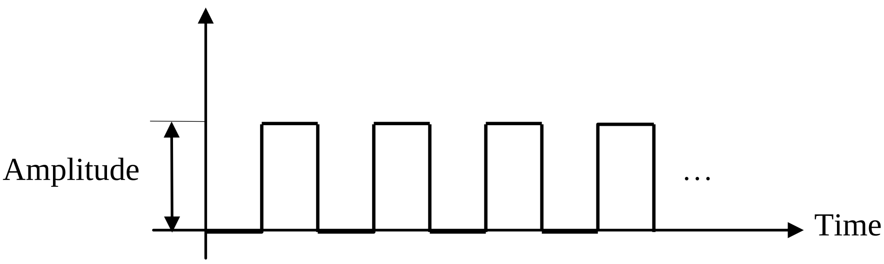


Amplitude, Period and Phase

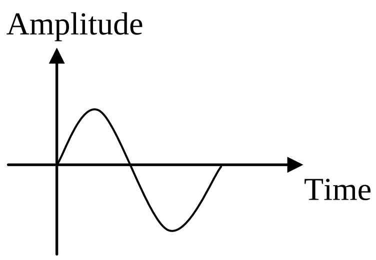
The three characteristics of periodic signals can be redefined for a periodic digital signal



No phase shift

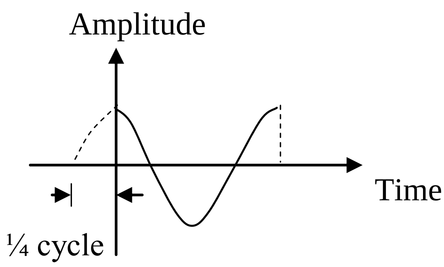


180^o phase shift

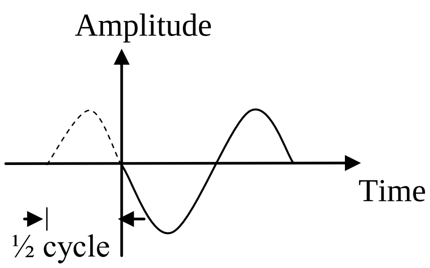


0^o phase shift

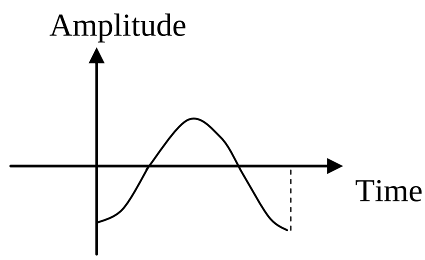
(no phase shift)



90^o phase shift



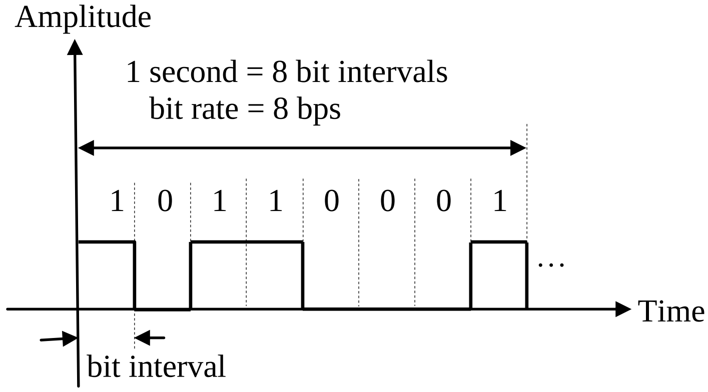
180^o phase shift



270^o phase shift

Bit Interval and Bit Rate

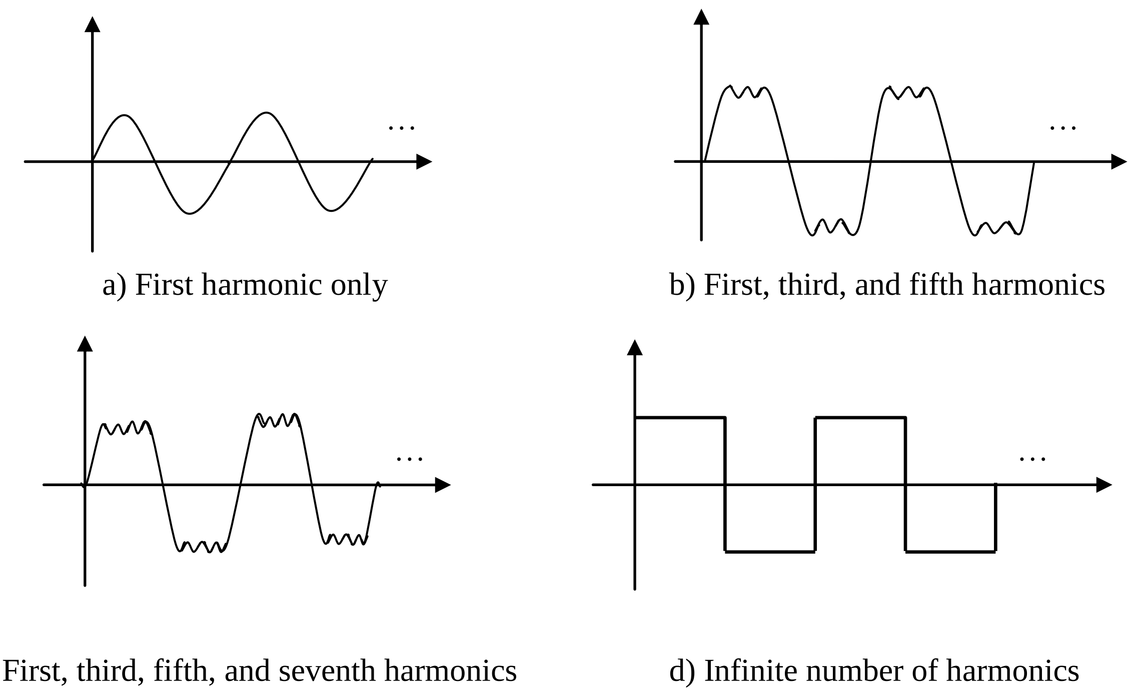
Most digital signals are aperiodic and thus terms like period or frequency are not appropriate. Two new terms, bit interval (instead of period) and bit rate (instead of frequency) are used to describe a digital signal. The bit interval is the time required to send one single bit. The bit rate is the number of bit intervals per second. This means that the bit rate is the number of bits sent in one second, usually expressed in bits per second (bps).



Decomposition of a Digital Signal

A digital signal can be decomposed into an infinite number of simple sine waves called harmonics, each with different amplitude, frequency and phase. This means that when a digital signal is sent along a transmission medium, an infinite number of simple signals is being sent.

Harmonics of a Digital Signal



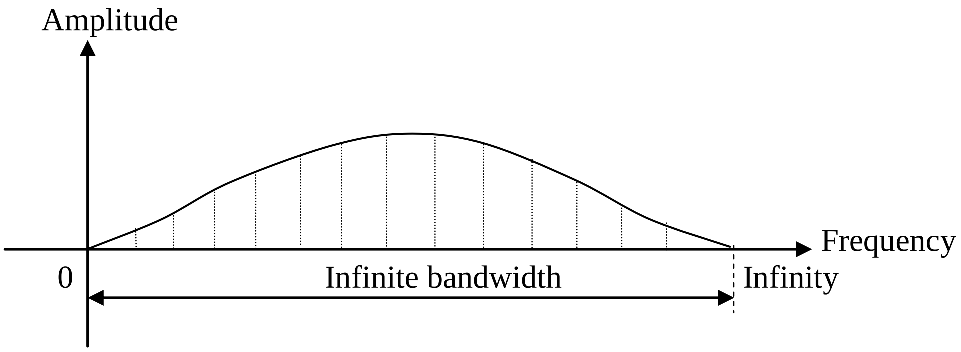
a) First harmonic only

b) First, third, and fifth harmonics

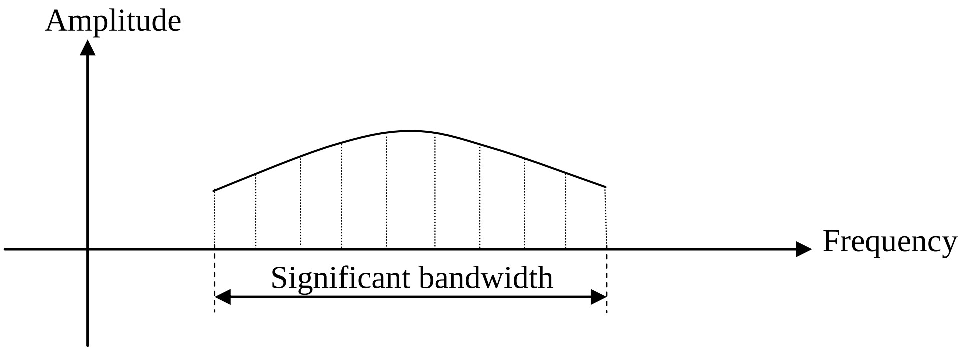
c) First, third, fifth, and seventh harmonics

d) Infinite number of harmonics

If some of the components do not pass through the medium, this results in distortion of the signal at the receiver side. Since no practical medium (such as a cable) is capable of transferring the entire range of frequencies, there will always be distortion.



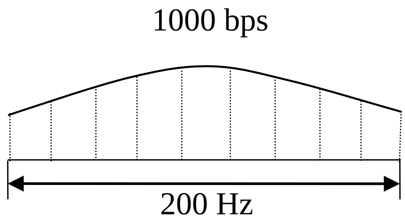
a) Spectrum for exact replica

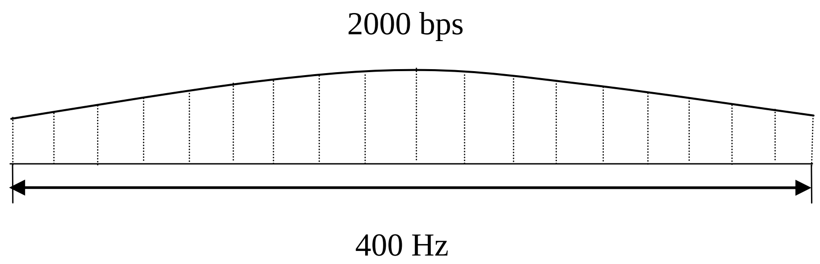


b) Significant spectrum

The part of the infinite spectrum whose amplitudes are significant (above an acceptable threshold), is called the significant spectrum, and its bandwidth is called the significant bandwidth.

When the bit rate increases, the significant bandwidth widens. For example, if the bit rate is 1000 bps, the significant bandwidth can be around 200 Hz, depending on the level of noise in the system. If the bit rate is 2000 bps, the significant bandwidth can be 400 Hz.

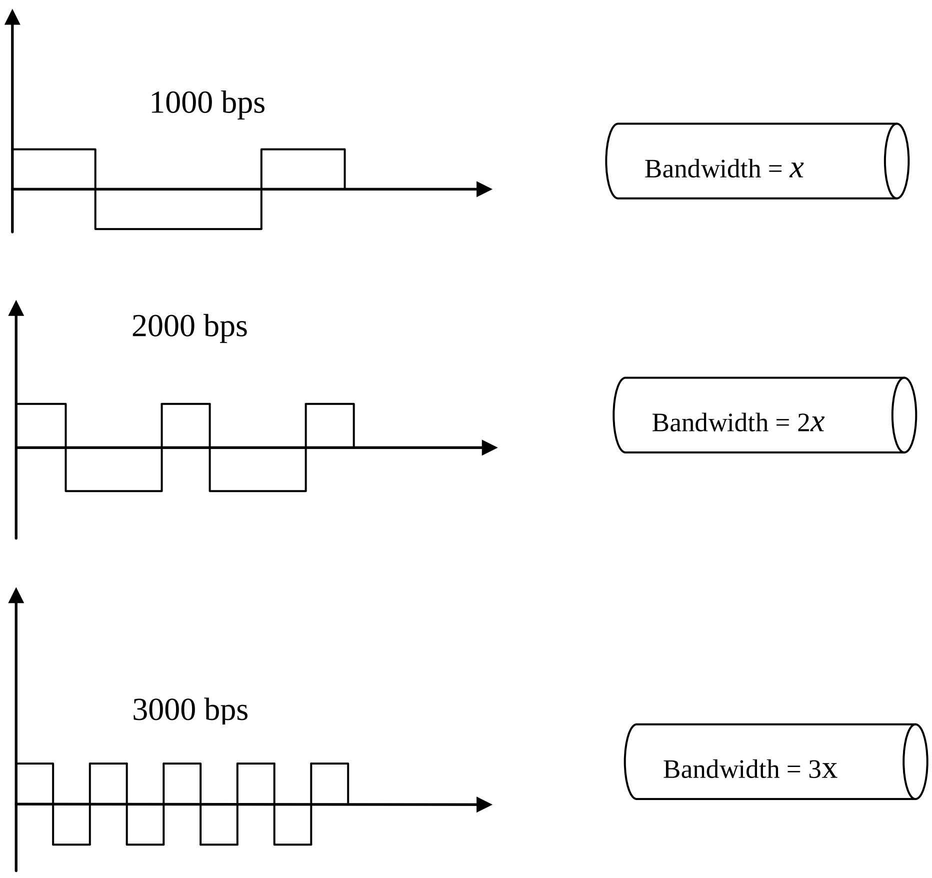




A transmission medium with a particular bandwidth is capable of transmitting only digital signals whose significant bandwidth is less than the bandwidth of the medium.

Channel Capacity

The maximum bit rate a transmission medium can transfer is called channel capacity of the medium. The capacity of a channel, however, depends on the type of encoding technique and the signal-to-noise ratio of the system. For example a normal telephone line with a bandwidth of 3000 Hz is capable of transferring up to 20,000 bps, but other factors, like noise, can decrease this rate.



Noise

In the absence of a signal, a transmission medium ideally has no electrical signal present. In practice, however, there is what we call line noise level, because of random perturbations on the line even when no signal is being transmitted. An important parameter associated with a transmission medium, therefore, is the ratio of the average power in a received signal, S, to the power in the noise level, N. The ratio S/N is known as the **signal-to-noise ratio (SNR)** and normally is expressed in decibels, that is:

$$\text{SNR} = 10\ \log_{10}\left(\frac{S}{N}\right) dB$$

- A high SNR ratio means a good-quality signal.
- A low SNR ratio means a low-quality signal.

The theoretical maximum data rate of transmission channel is related to the SNR ratio and we can determine this rate using a formula attributed to Shannon and Hartley. This is known as the **Shannon-Hartley Law**, which states:

$$C = W \log_2\left(1+\frac{S}{N}\right)\text{bps}$$

$$\approx 3.32\ W \log_{10}\left(1+\frac{S}{N}\right)\text{bps}$$

where C is the data rate in bps, W is the bandwidth of the line channel in Hz, S is the average signal power in watts and N is the random noise power in watts.

Ex

Consider a voice channel with BW of 2,800 Hz. A typical value of S/N for a telephone line is 20 dB. What is the channel capacity?

Solution

SNR = 20 dB

20 = 10 $\log_{10}$ (S/N) $\Rightarrow$ S/N = 100

W = 2,800 Hz

$$C = W \log_2\left(1+\frac{S}{N}\right)\text{bps} \quad \approx 3.32\ W \log_{10}\left(1+\frac{S}{N}\right)\text{bps}$$

$$\approx 3.32\ (2800) \log_{10}(1+100)$$

$$C = 18{,}632 \text{ bps}$$

CHAPTER 3

TRANSMISSION MEDIA

There are two basic categories of transmission media: guided and unguided media.

Guided transmission media use cabling system that guides the data signals along a specific path. Data signals are bound by the cabling system. Guided media is also known as bound media. “Cabling” is meant in a generic sense, and is not meant to be interpreted as copper wire cabling only.

Unguided transmission media consists of a means for the data signals to travel but nothing to guide them along a specific path. The data signals are not bound to a cabling media and are therefore often called unbound media.

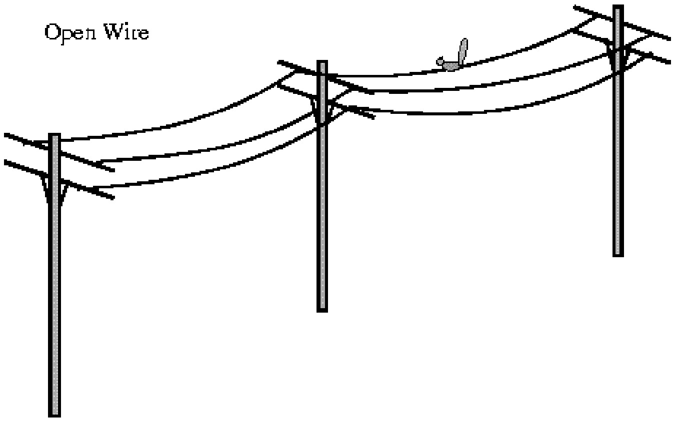
Transmission Media: Guided

There four basic types of guided media:

a. Open Wire
b. Twisted Pair
c. Coaxial Cable
d. Optical Fibre

Open Wire

Open wire is traditionally used to describe the electrical wire strung along power poles. There is a single wire strung between poles. No shielding or protection from noise interference is used. We are going to extend the traditional definition of open wire to include any data signal path without shielding or protection from noise interference. This can include multi conductor cables or single wires. This medium is susceptible to a large degree of noise and interference and consequently is not acceptable for data transmission except for short distances under 20 ft.

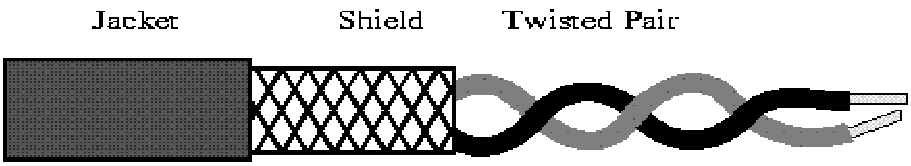


Twisted Pair

The wires in twisted pair cabling are twisted together in pairs. Each pair consists of a wire used for the positive data signal and a wire used for the negative data signal. Any noise that appears on one wire of the pair will also occur on the other wire. Since the wires have opposite polarities, they are 180 degrees out of phase. When noise appears on both wires, it cancels or nulls itself out at the receiving end. Twisted pair cables are most effectively used in systems that use a balanced line method of transmission: polar line coding (Manchester encoding) as opposed to unipolar line coding.

Unshielded Twisted Pair

The degree of reduction in noise interference is determined specifically by the number of turns per foot. Increasing the number of turns per foot reduces the noise interference. To further improve noise rejection, a foil or wire braid “shield” is woven around the twisted pairs. This shield can be woven around individual pairs or around a multi-pair conductor (several pairs).



Shielded Twisted Pair

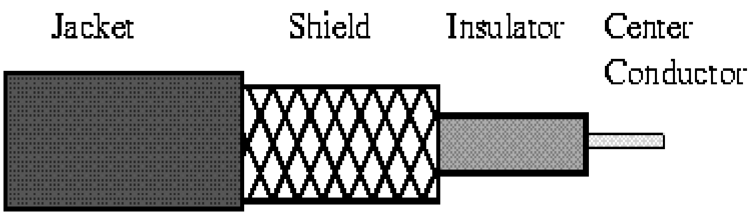
Cables with a shield are called shielded twisted pair and are commonly abbreviated STP. Cables without a shield are called unshielded twisted pair or UTP. Twisting the wires together results in a characteristic impedance for the cable. Typical impedance for UTP is 100 Ohm for Ethernet 10BaseT cable.

UTP or unshielded twisted pair cable is used in Ethernet 10BaseT and can also be used with Token Ring. It uses the RJ line of connectors (RJ45, RJ11, etc..).

STP or shielded twisted pair is used with the traditional Token Ring cabling or ICS-IBM Cabling System. It requires a custom connector. IBM STP (shielded twisted pair) has a characteristic impedance of 150 Ohm.

Coaxial Cable

Coaxial cable consists of two conductors. The inner conductor is held inside an insulator with the other conductor woven around it providing a shield. An insulating protective coating called a jacket covers the outer conductor.



Coaxial Cable

The outer shield protects the inner conductor from outside electrical signals. The distance between the outer conductor (shield) and inner conductor, plus the type of material used for insulating the inner conductor determine the cable properties or impedance. Typical impedances for coaxial cables are 75 Ohms for TV cable, 50 Ohms for Ethernet Thinnet and Thicknet. The excellent control of the impedance characteristics of the cable allow higher data rates to be transferred than with twisted pair cable.

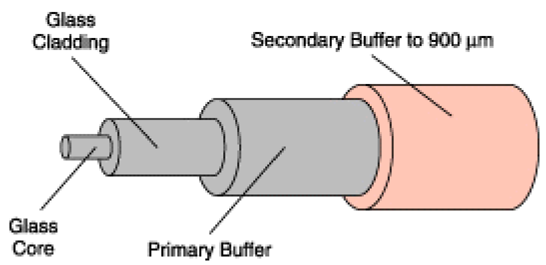
Optical fibre

Optical fibre consists of thin glass fibres that can carry information at frequencies in the visible light spectrum and beyond. The typical optical fibre consists of a very narrow strand of glass called the core. Around the core is a concentric layer of glass called the cladding. A typical core diameter is 62.5 microns (1 micron = 10^{-6} m). Typically Cladding has a diameter of 125 microns. Coating the cladding is a protective coating consisting of plastic, it is called the Jacket.

Fibre Optic Cables

Just as standard electric cables come in a variety of sizes, shapes, and types, fibre optic cables are available in different configurations. The simplest cable is just a single strand of fibre, whereas complex cables are made up of multiple fibres with different layers and other elements.

The portion of a fibre optic cable (core) that carries the light is made from either glass or plastic. Another name for glass is silica. Special techniques have been developed to create nearly perfect optical glass or plastic, which is transparent to light. Such materials can carry light over a long distance. Glass has superior optical characteristics over plastic. However, glass is far more expensive and more fragile than plastic. Although the plastic is less expensive and more flexible, its attenuation of light is greater. For a given intensity, light will travel a greater distance in glass than in plastic. For very long distance transmission, glass is certainly preferred. For shorter distances, plastic is much more practical.



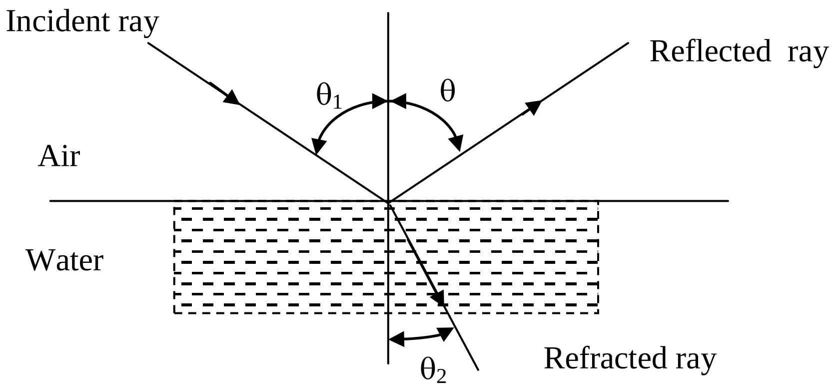
All fibres consist of a number of substructures including:

A core, which carries most of the light, surrounded by
A cladding, which bends the light and confines it to the core, surrounded by
A substrate layer (in some fibres) of glass which does not carry light, but adds to the diameter and strength of the fibre, covered by
A primary buffer coating, this provides the first layer of mechanical protection, covered by
A secondary buffer coating, this protects the relatively fragile primary coating.

The cladding is also made of glass or plastic but has a lower index of refraction. This ensures that the proper interface is achieved so that the light waves remain within the core. In addition to protecting the fibre core from nicks and scratches, the cladding adds strength. Some fibre optic cables have a glass core with a glass cladding. Others have a plastic core with a plastic cladding. Another common arrangement is a glass core with a plastic cladding. It is called plastic-clad silica (PCS) cable.

An important characteristic of fibre optics is refraction. Refraction is the characteristic of a material to either pass or reflect light. When light passes through a medium, it "bends" as it passes from one medium to the other. An example of this is when we look into a pond of water.

In 1621, the Dutch mathematician Willebrard Snell established that rays of light can be traced as they propagate from one medium to another based on their indices of refraction. Snell's low is stated by the equation:



$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1};$$

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

where n_1-refractive index of material 1; θ_1-angle of incidence; θ_2-angle of refraction; n_2-refractive index of material 2. When the angle of incidence, θ_1, becomes large enough to cause the sine of the refraction angle, θ_2, to exceed the value of 1, total internal reflection occurs. This angle is called the critical angle, θ_c. The critical angle, θ_c, can be derived from Snell's law as follows

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

$$\sin \theta_1 = n_2 \sin \theta_2 / n_1$$

When $\sin \theta_1 = \sin \theta_2$, then $\sin \theta_1 = n_2 / n_1$. Therefore, the critical angle: $\theta_c = \sin^{-1} (n_2 / n_1)$
Its index of refraction, however, it is typically 1% less than that of its core. This permits total internal reflection of rays entering the fibre and striking the core-cladding interface above the critical angle of approximately 82-degree ($\sin^{-1}$ (1/1.01). The core of the fibre therefore guides the light and the cladding contains the light. The cladding material is much less transparent than the glass making up the core of the fibre. This causes light rays to be absorbed if they strike the core-cladding interface at an angle less than the critical angle.

If the angle of incidence is small, the light rays are reflected and do not pass into the water. If the angle of incident is great, light passes through the media but is bent or refracted.

In the following figure, a light ray is transmitted into the core of an optical fibre. Total

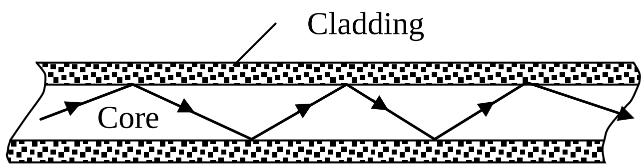


Figure 2

internal reflection occurs as it strikes the lower index cladding material.

Optical fibres work on the principle that the core refracts the light and the cladding reflects the light. The core refracts the light and guides the light along its path. The cladding reflects any light back into the core and stops light from escaping through it.

Transmission Modes

There are three primary types of transmission modes using optical fibre. They are

a. Step Index
b. Graded Index
c. Single Mode

Step index has a large core, so the light rays tend to bounce around inside the core, reflecting off the cladding. This causes some rays to take a longer or shorter path through the core. Some take the direct path with hardly any reflections while others bounce back and forth taking a longer path. The result is that the light rays arrive at the receiver at different times. The signal becomes longer than the original signal. LED light sources are used. Typical Core: 62.5 microns.

Graded index has a gradual change in the core's refractive index. This causes the light rays to be gradually bent back into the core path. This is represented by a curved reflective path in the attached drawing. The result is a better receive signal than with step index. LED light sources are used. Typical Core: 62.5 microns.

Note: Both step index and graded index allow more than one light source to be used (different colours simultaneously), so multiple channels of data can be run at the same time!

Single mode has separate distinct refractive indexes for the cladding and core. The light ray passes through the core with relatively few reflections off the cladding. Single mode is used for a single source of light (one colour) operation. It requires a laser and the core is very small: 9 microns.

Basic Construction of Fibre-Optic Cables

There are two basic ways of classifying fibre optic cables. The first way is an indication of how the index of refraction varies across the cross section of the cable. The second way of classification is by mode. Mode refers to the various paths that the light rays can take in passing through the fibre. Usually these two methods of classification are combined to define the types of cable. There are two basic ways of defining the index of refraction variation across a cable. These are step index and graded index. Step index refers to the fact that there is a sharply defined step in the index of refraction where the fibre core and the cladding interface. It means that the core has one constant index of refraction n_1, while the cladding has another constant index of refraction n_2.

The other type of cable has a graded index. In this type of cable, the index of refraction of the core is not constant. Instead, the index of refraction varies smoothly and continuously over the diameter of the core. As you get closer to the centre of the core, the index of refraction gradually increases, reaching a peak at the centre and then declining as the other outer edge of the core is reached. The index of refraction of the cladding is constant.

Mode refers to the number of paths for the light rays in the cable. There are two classifications: single mode and multimode. In single mode, light follows a single path through the core. In multimode, the light takes many paths through the core.

Each type of fibre optic cable is classified by one of these methods of rating the index or mode. In practice, there are three commonly used types of fibre optic cable: multimode step index, single mode step index and multimode graded index cables.

1. Multimode Step-Index Fibre.

This cable (see Figure 5 (a)) is the most common and widely used type. It is also the easiest to make and, therefore, the least expensive. It is widely used for short to medium distances at relatively low pulse frequencies.

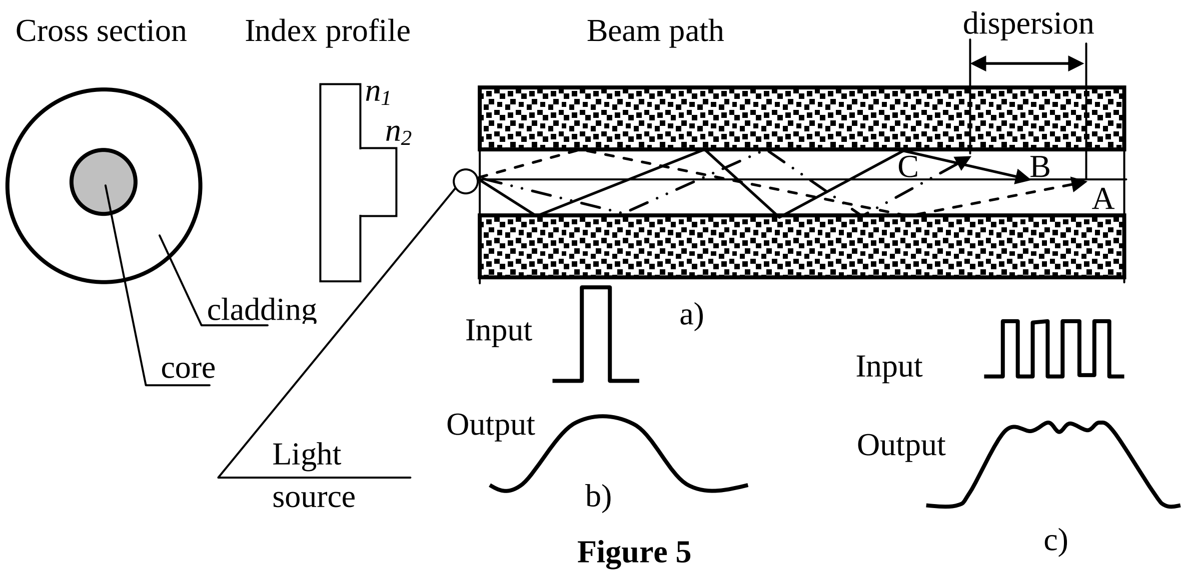


Figure 5

The main advantage of a multimode step index fibre is the large size. Typical core diameters are in the 50-to-1000 micrometers (μm) range. Such large diameter cores are excellent at gathering light and transmitting it efficiently. This means that an inexpensive light source such as LED can be used to produce the light pulses. The light takes many hundreds of even thousands of paths through the core before exiting. Because of the different lengths of these paths, some of the light rays take longer to reach the end of the cable than others. The problem with this is that it stretches the light pulses (Figure 5 (b). In Figure 5 ray A reaches the end first, then B, and C. The result is a pulse at the other end of the cable that is lower in amplitude due to the attenuation of the light in the cable and increased in duration due to the different arrival times of the various light rays. The stretching of the pulse is referred to as modal dispersion. Because the pulse has been stretched, input pulses can not occur at a rate faster than the output pulse duration permits. Otherwise the pulses will essentially merge together as shown in Figure 5 (c). At the output, one long pulse will occur and will be indistinguishable from the three separate pulses originally transmitted. This means that incorrect information will be received. The only core for this problem is to reduce the pulse repetition rate. When this is done, proper operation occurs. But with pulses at a lower frequency, less information can be handled.

2. Single Mode Cable

In a single mode, or mono-mode, step-index fibre cable the core is so small that the total number modes or paths through the core are minimised and modal dispersion is essentially eliminated. The typical core sizes are 5 to 15 μm. The output pulse has essentially the same duration as the input pulse (see Figure 6).

The single mode step index fibres are by far the best since the pulse repetition rate can be high and the maximum amount of information can be carried. For very long distance transmission and maximum information content, single-mode step-index fibre cables should be used.

The main problem with this type of cable is that because of its extremely small size, it is difficult to make and is, therefore, very expensive. Handling, splicing, and making interconnections are also more difficult. Finally, for proper operation an expensive, super intense light source such as a laser must be used. For long distances, however, this is the type of cable preferred.

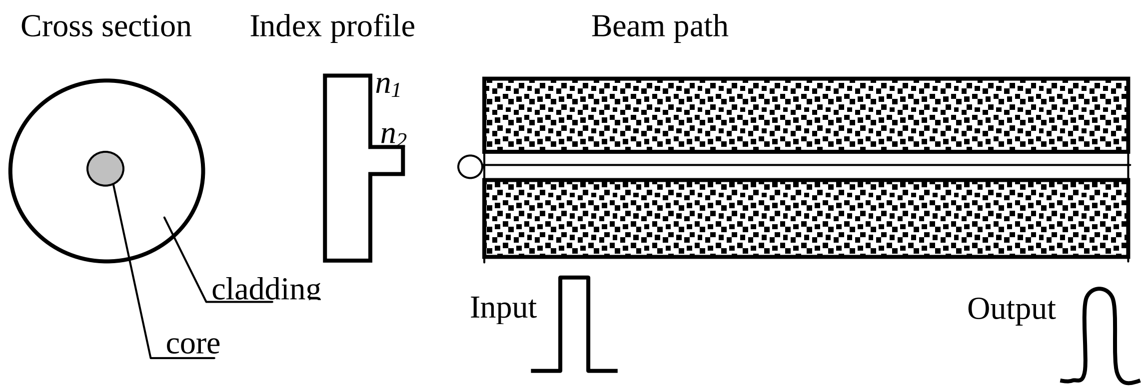


Figure 6

3. Multimode graded-index fibre cables

These cables have a several modes or paths of transmission through the cable, but they are much more orderly and predictable. Figure 7 shows the typical paths of the light beams. Because of the continuously varying index of refraction across the core, the light rays are bent smoothly and converge repeatedly at points along the cable.

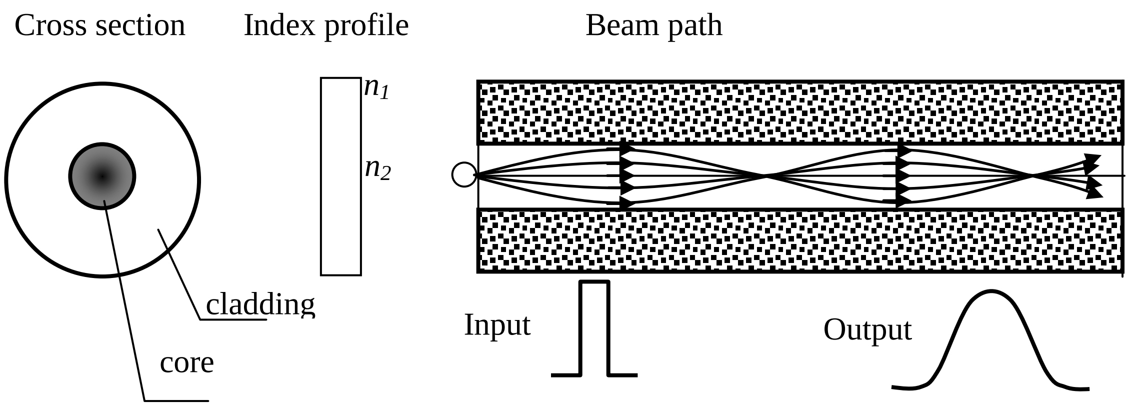


Figure 7

The light rays near the edge of the core take a longer path but travel faster since the index of refraction is lower. All the modes or light paths tend to arrive at one point simultaneously. The result is that there is less modal dispersion. It is not eliminated entirely, but the output pulse is not nearly as stretched as in multimode step index cable. The output pulse is only slightly elongated. As a result, this cable can be used at very high pulse rates and, therefore, a considerable amount of information can be carried on it.

This type of cable is also much wider in diameter with core sizes in the 50 to 100 (μm) range. Therefore, it is easier to splice and interconnect, cheaper, and less-intense light sources may be used. The most popular fibre-optic cables that are used in LAN: multimode-step index cable -65.5/125; multimode-graded index cable - 50/125. The multimode-graded index cable - 100/140 or 200/300 are recommended for industrial control applications because of its large size. In high data rate systems single mode fibre 9/125 is used. Typical core and cladding diameters of these cables are shown in Figure 8.

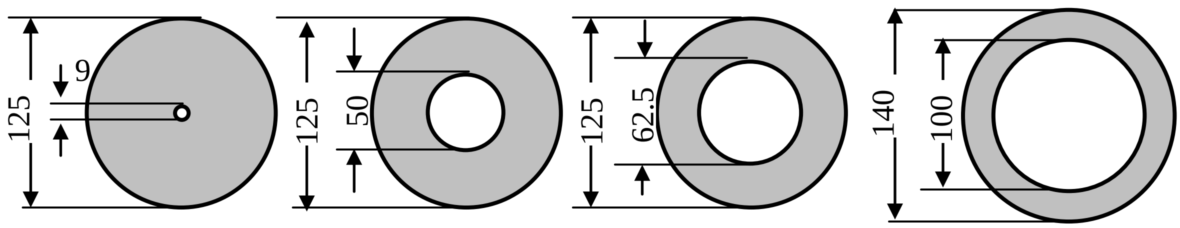


Figure 8

Specifications of the Fibre Cables

Indoor cable specifications:

- LED (Light Emitting Diode) light source
- 3.5 dB/Km Attenuation (loses 3.5 dB of signal per kilometer)

- 850 nM - wavelength of light source
- Typically 62.5/125 (core diameter/cladding diameter)
- Multimode - can run many light sources.

Outdoor cable specifications:

- Laser Light Source
- 1 dB/Km Attenuation (loses 1 dB of signal per kilometer)
- 1170 nM - wavelength of light source
- Monomode (single mode)

Advantages of Optical Fibre:

- Noise immunity: RFI and EMI immune (RFI - Radio Frequency Interference, EMI - Electromagnetic Interference)
- Security: cannot tap into cable.
- Large Capacity due to BW (bandwidth)
- No corrosion
- Longer distances than copper wire
- Smaller and lighter than copper wire
- Faster transmission rate

Disadvantages of optical fibre:

- Physical vibration will show up as signal noise!
- Limited physical arc of cable. Bend it too much and it will break!
- Difficult to splice

The cost of optical fibre is a trade-off between capacity and cost. At higher transmission capacity, it is cheaper than copper. At lower transmission capacity, it is more expensive.

Media versus Bandwidth

The following table compares the usable bandwidth of the different guided transmission media.

Cable Type	Bandwidth
Open Cable	0 - 5 MHz
Twisted Pair	0 - 100 MHz
Coaxial Cable	0 - 600 MHz
Optical Fibre	0 - 1 GHz

Transmission Media: Unguided

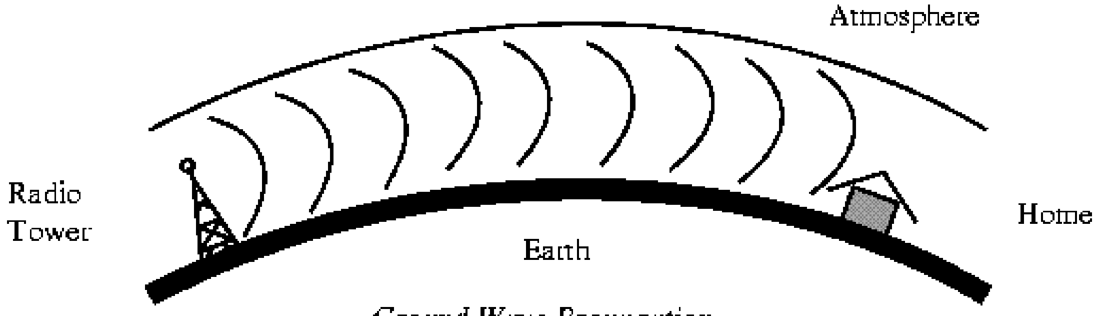
Unguided transmission media is data signals that flow through the air. They are not guided or bound to a channel to follow. They are classified by the type of wave propagation.

RF Propagation

There are three types of RF (radio frequency) propagation:

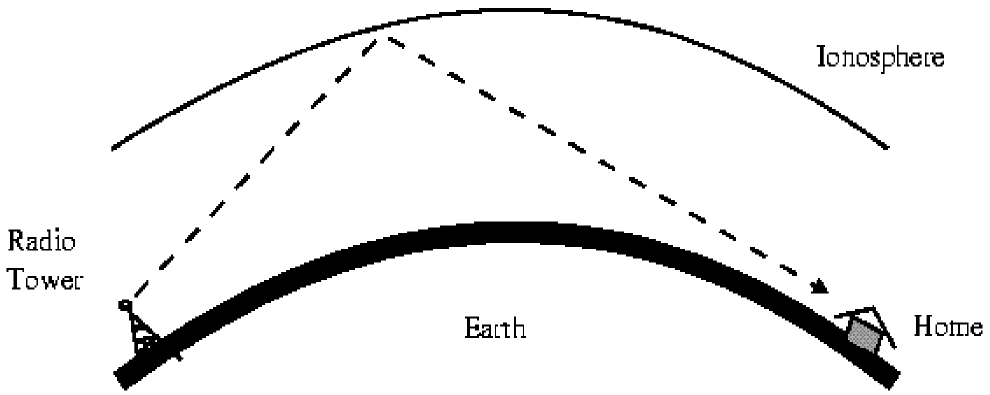
- Ground Wave
- Sky Wave
- Line of Sight (LOS)

Ground wave propagation follows the curvature of the Earth. Ground waves have carrier frequencies up to 2 MHz. AM radio is an example of ground wave propagation.

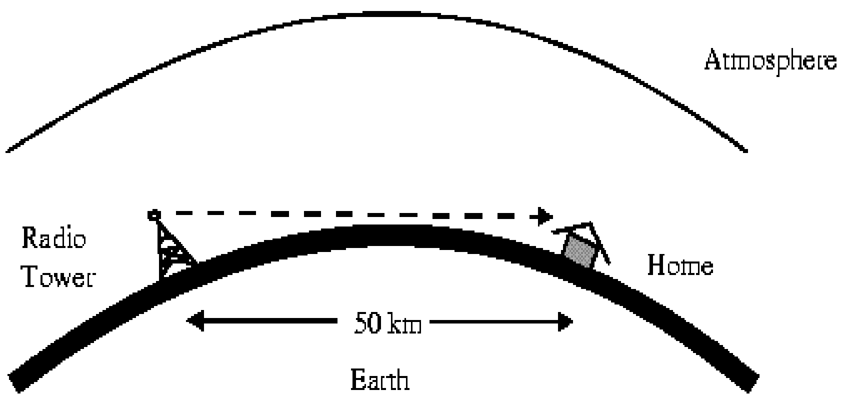


Ground Wave Propagation

Sky wave propagation bounces off of the Earth's ionospheric layer in the upper atmosphere. It is sometimes called double hop propagation. It operates in the frequency range of 30-85 MHz. Because it depends on the Earth's ionosphere, it changes with the weather and time of day. The signal bounces off of the ionosphere and back to earth. Ham radios operate in this range.



Line of sight propagation transmits exactly in the line of sight. The receive station must be in the view of the transmit station. It is sometimes called space waves or troposphere propagation. It is limited by the curvature of the Earth for ground-based stations (100 km, from horizon to horizon). Reflected waves can cause problems. Examples of line of sight propagation are: FM radio, microwave and satellite.



Radio Frequencies

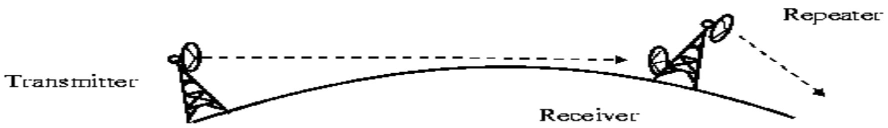
The frequency spectrum operates from 0 Hz (DC) to gamma rays (10^{19} Hz).

Name	Frequency (Hertz)	Examples
Gamma Rays	10^{19}	
X-Rays	10^{17}	
Ultra-Violet Light	7.5×10^{15}	
Visible Light	4.3×10^{14}	
Infrared Light	3×10^{11}	
EHF - Extremely High Frequencies	30 GHz (Giga = 10^9)	Radar
SHF - Super High Frequencies	3 GHz	Satellite & Microwaves
UHF - Ultra High Frequencies	300 MHz (Mega = 10^6)	UHF TV (Ch. 14-83)
VHF - Very High Frequencies	30 MHz	FM & TV (Ch2 - 13)
HF - High Frequencies	3 MHz2	Short Wave Radio
MF - Medium Frequencies	300 kHz (kilo = 10^3)	AM Radio
LF – Low Frequencies	30 kHz	Navigation
VLF - Very Low Frequencies	3 kHz	Submarine Communications
VF - Voice Frequencies	300 Hz	Audio
ELF - Extremely Low Frequencies	30 Hz	Power Transmission

Radio frequencies are in the range of 300 kHz to 10 GHz. We are seeing an emerging technology called wireless LANs. Some use radio frequencies to connect the workstations together, some use infrared technology.

Microwave

Microwave transmission is line of sight transmission. The transmit station must be in visible contact with the receive station. This sets a limit on the distance between stations depending on the local geography. Typically the line of sight due to the Earth's curvature is only 100 km to the horizon! Repeater stations must be placed so the data signal can hop, skip and jump across the country.



Microwaves operate at high operating frequencies of 3 to 10 GHz. This allows them to carry large quantities of data due to their large bandwidth.

Advantages:

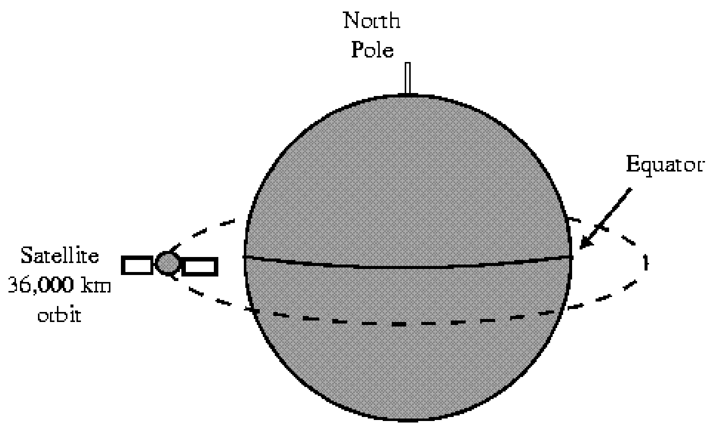
a. They require no right of way acquisition between towers.
b. They can carry high quantities of information due to their high operating frequencies.
c. Low cost land purchase: each tower occupies only a small area.
d. High frequency/short wavelength signals require small antennae.

Disadvantages:

a. Attenuation by solid objects: birds, rain, snow and fog.
b. Reflected from flat surfaces like water and metal.
c. Diffracted (split) around solid objects.
d. Refracted by atmosphere, thus causing beam to be projected away from receiver.

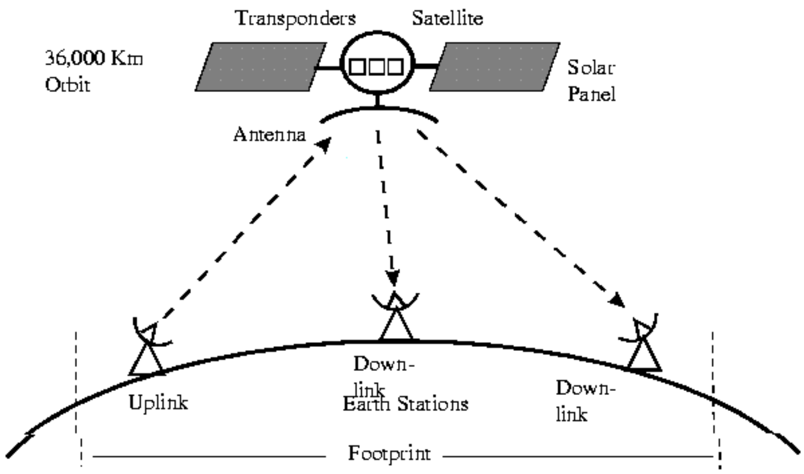
Satellite

Satellites are transponders (units that receive on one frequency and retransmit on another) that are set in geostationary orbits directly over the equator. These geostationary orbits are 36,000 km from the Earth's surface. At this point, the gravitational pull of the Earth and the centrifugal force of Earth's rotation are balanced and cancel each other out. Centrifugal force is the rotational force placed on the satellite that wants to fling it out into space.



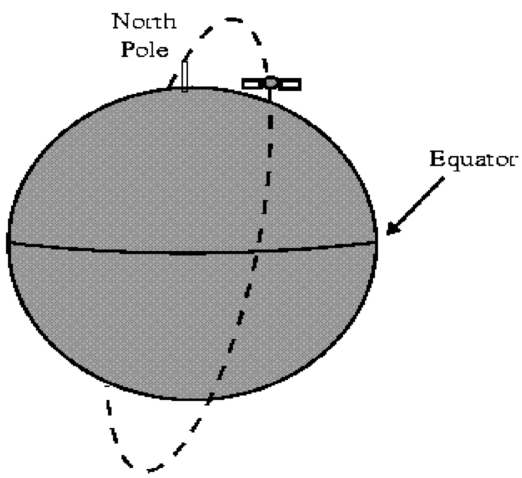
The uplink is the transmitter of data to the satellite. The downlink is the receiver of data. Uplinks and downlinks are also called Earth stations because they are located on the Earth.

The footprint is the "shadow" that the satellite can transmit to, the shadow being the area that can receive the satellite's transmitted signal**.**



Iridium Telecom System

The Iridium Telecom System is a new satellite system that will be the largest private aerospace project. It is a mobile telecom system intended to compete with cellular phones. It relies on satellites in lower Earth orbit (LEO). The satellites will orbit at an altitude of 900 - 10,000 km in a polar, non-stationary orbit. Sixty-six satellites are planned. The user's handset will require less power and will be cheaper than cellular phones. There will be 100% coverage of the Earth.



Unfortunately, although the Iridium project was planned for 1996-1998, with 1.5 million subscribers by end of the decade, it looked very financially unstable.

CHAPTER 4

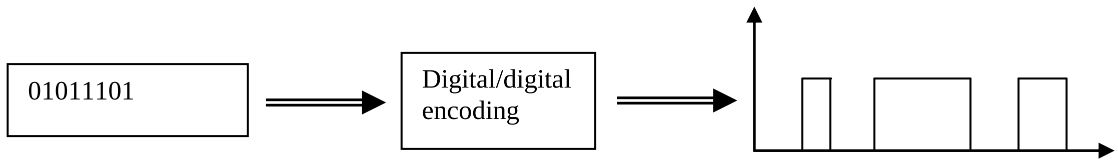
ENCODING, MODULATING & TRANSMISSION CODES

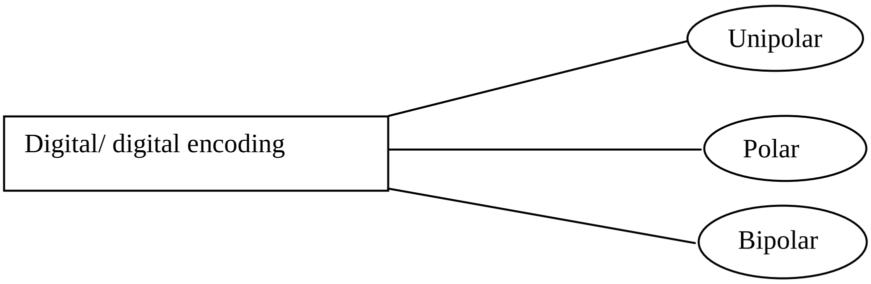
ENCODING (D/D) (A/D) (D/A) (A/A)

Information must be encoded into signals before it can be transported across communication media. We must encode data into signals to send them from one place to another.

Digital-to-Digital Encoding

Digital-to-Digital Encoding is the representation of digital information by a digital signal. (eg. computer-to-printer)



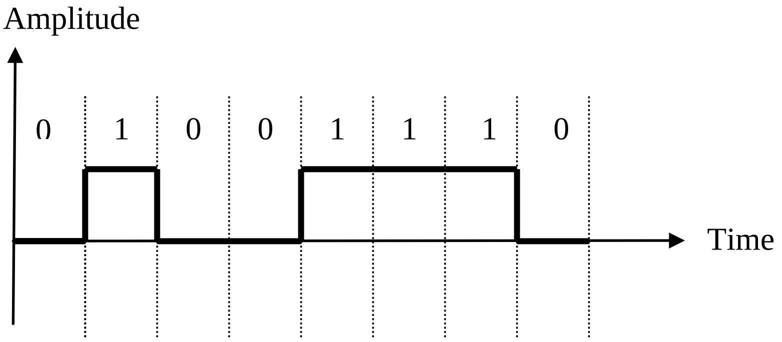


Unipolar

Digital transmission systems work by sending voltage pulses along a media link, usually a wire or a cable. In most types of encoding, one voltage level stands for binary 0 and another level stands for binary 1. The polarity of a pulse refers to whether it is positive or negative.

Unipolar encoding is so named because it uses only one polarity. Therefore, only one of the two binary states is encoded, usually the 1. The other state, usually 0, is represented by zero voltage, or an idle line.

Unipolar encoding uses only one level of value.



1's encoded as positive, 0's are idle. Unipolar encoding is straight forward and inexpensive to implement. However, it has two problems that make it unusable: DC component and synchronisation.

DC component

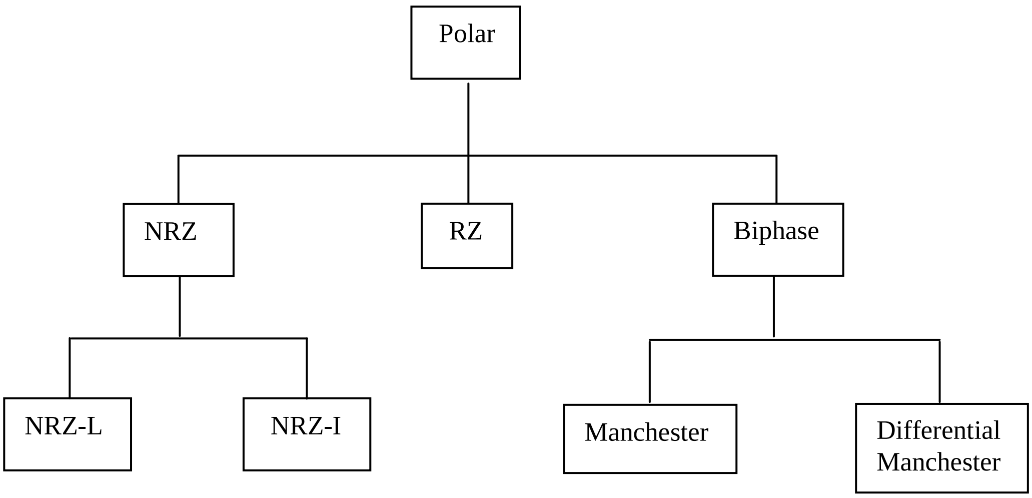
Average amplitude is nonzero → creates a direct current (DC) component, when a signal contains a DC component it cannot travel through media that cannot handle DC components: e.g. microwaves or transformers.

Synchronisation

When a signal is unvarying, the receiver cannot determine the beginning and ending of each bit. Therefore, Synchronisation problem in unipolar encoding can occur whenever the data stream includes a long uninterrupted series of 1's or 0's.

<u>Polar Encoding</u>

Polar encoding uses two voltage levels: one positive and one negative. In most polar encoding methods the average voltage level on the line is reduced and the DC component problem of unipolar encoding is alleviated.



Non-Return-to-Zero (NRZ) Encoding

In NRZ encoding, the level of the signal is always either positive or negative. In NRZ system if the line is idle it means no transmission is occurring at all.

- **NRZ-L (Non-return-to-zero, Level)**

In NRZ-L the level of the signal is dependant upon the state of the bit.

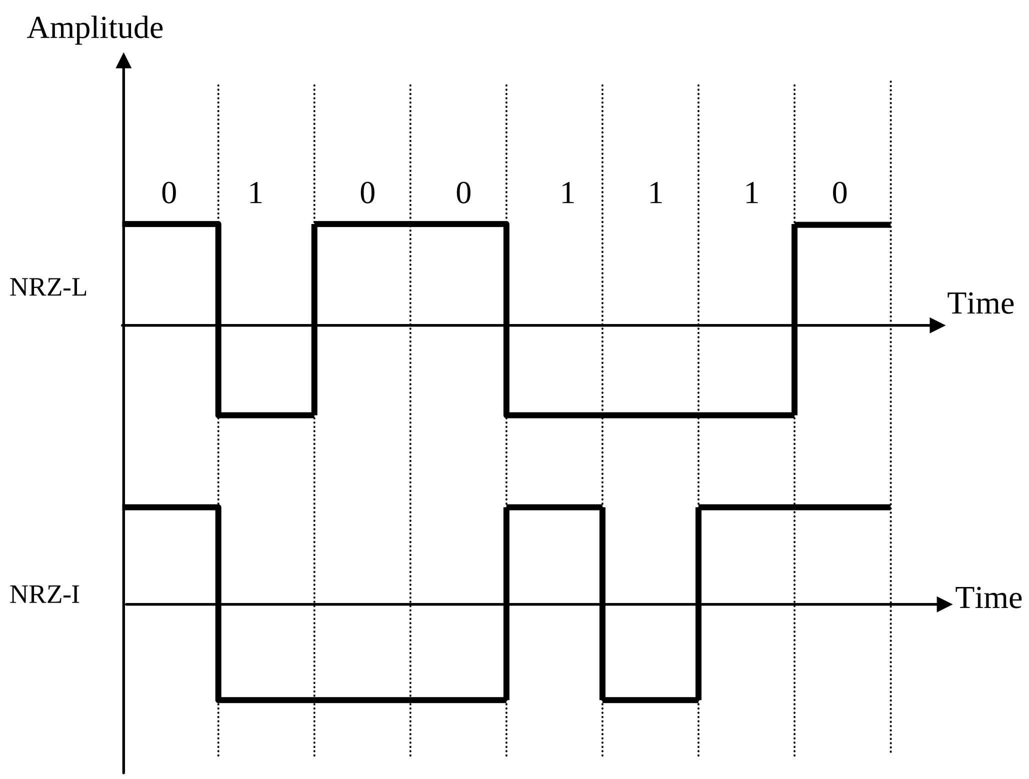
A positive voltage usually means the bit is 0, and negative voltage means the bit is a 1 (and vice versa).

- **NRZ-I (Non-return-to-zero, Invert)**

In NRZ-I, an inversion of the voltage level represents a 1 bit. It is the transition between a positive and a negative voltage, not the voltages themselves that represents a 1 bit. A 0 bit is represented by no change.

An advantage of NRZ-I over NRZ-L is that because the signal changes every time a 1 bit is encountered, it provides some synchronisation.

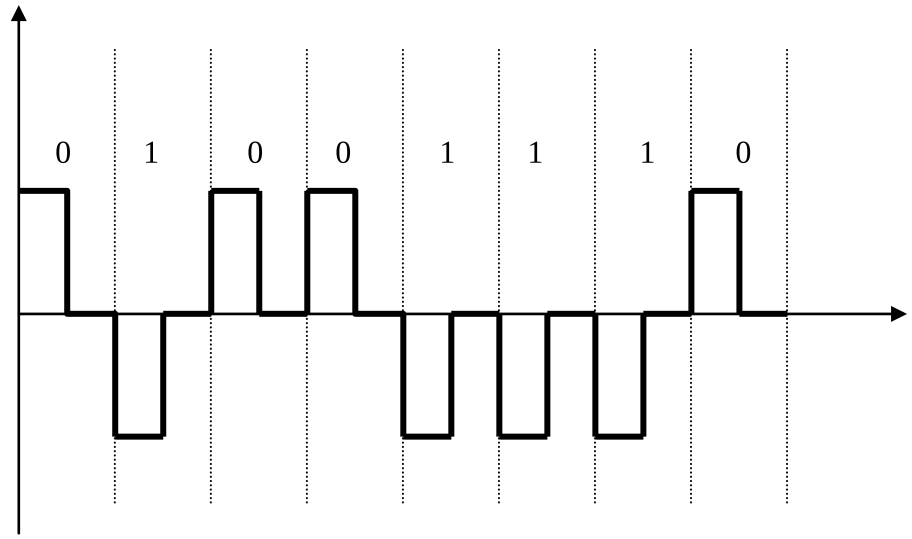
Each inversion allows the receiver to synchronise its timer to the actual arrival of the transmission.



RZ (Return-to-zero) Encoding

To assure synchronisation, there must be a signal change for each bit. The receiver can use these changes to built up, update, and synchronise its clock.

One solution is return to zero (RZ) encoding, which uses three Values: positive, negative, and zero.



The main disadvantage of RZ encoding is that it requires two signal changes to encode one bit and therefore occupies more bandwidth. But of the three alternatives discussed above, it is the most effective. Because a good encoded digital signal must contain a provision for synchronisation.

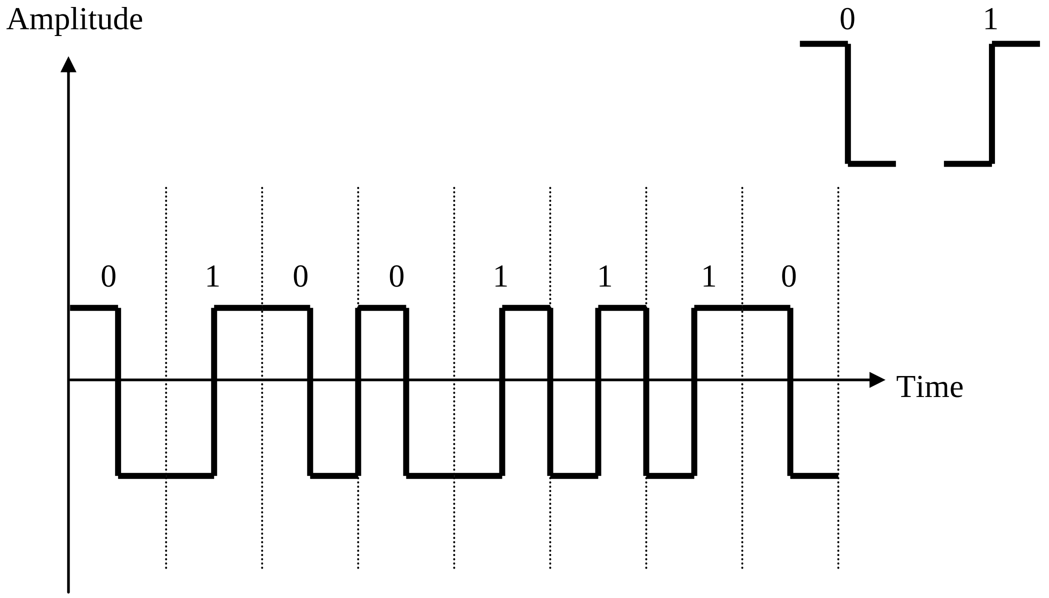
Biphase Encoding

Probably the best existing solution to the problem of synchronisation is biphase encoding. In this method, the signal changes at the middle of the bit interval but does not return to zero. Instead it continues to the opposite pole. As in RZ, these mid-interval transitions allow for synchronisation.

Biphase encoding is implemented in two different ways: Manchester and differential Manchester.

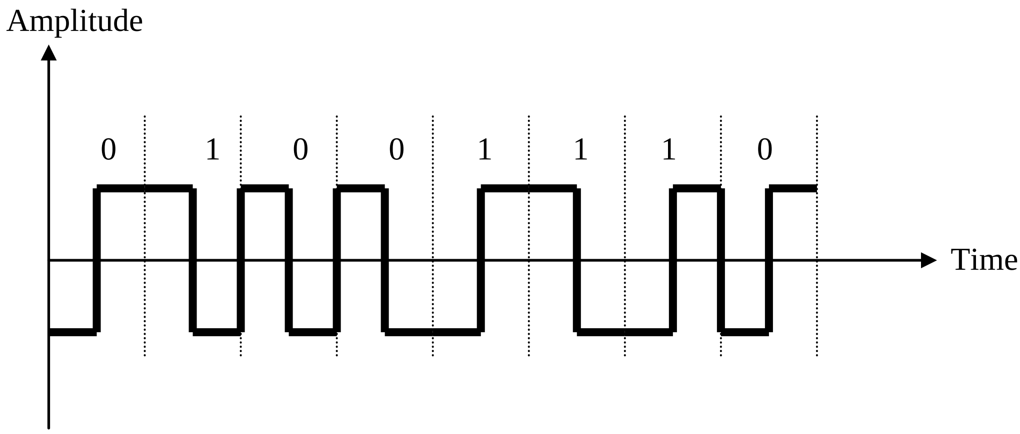
- **Manchester**

Manchester encoding uses the inversion at the middle of each bit interval for both synchronisation and bit representation. A negative-to-positive transition represents binary 1 and a positive-to-negative transition represents binary 0.



- **Differential Manchester**

  In this method, the inversion at the middle of the bit is used for synchronisation, but the presence or absence of an additional transition at the beginning of the interval is used to identify the bit. A transition means binary 0 and no transition means binary 1. The bit representation is shown by the inversion and non-inversion at the beginning of the bit.



Bipolar Encoding

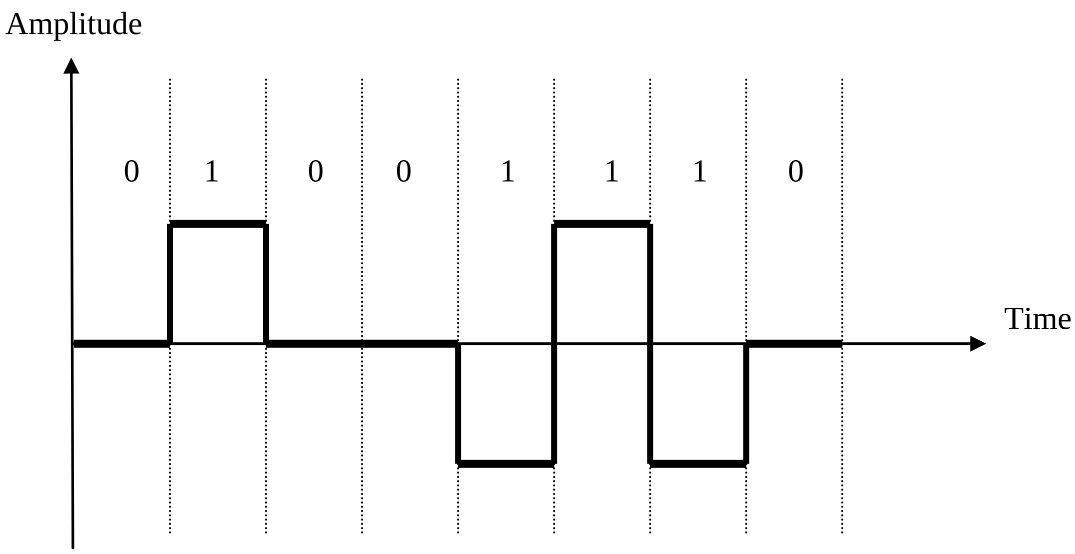
Bipolar encoding uses three voltage levels: positive, negative and zero. The zero level is used to represent binary 0 positive and negative voltages represent alternating 1s. (If 1^{st} one +ve, 2^{nd} is -ve).

* Three types of bipolar encoding are popular use by the data communications industry: AMI, B8ZS, and HDB3.

- **Bipolar Alternate Mark Inversion (AMI)**

  Bipolar AMI is the simplest type of bipolar encoding. The word mark comes from telegraphy and means 1.

  AMI means alternate 1 inversion. A neutral, zero voltage represents binary 0. Binary 1s are represented by alternating positive and negative voltages



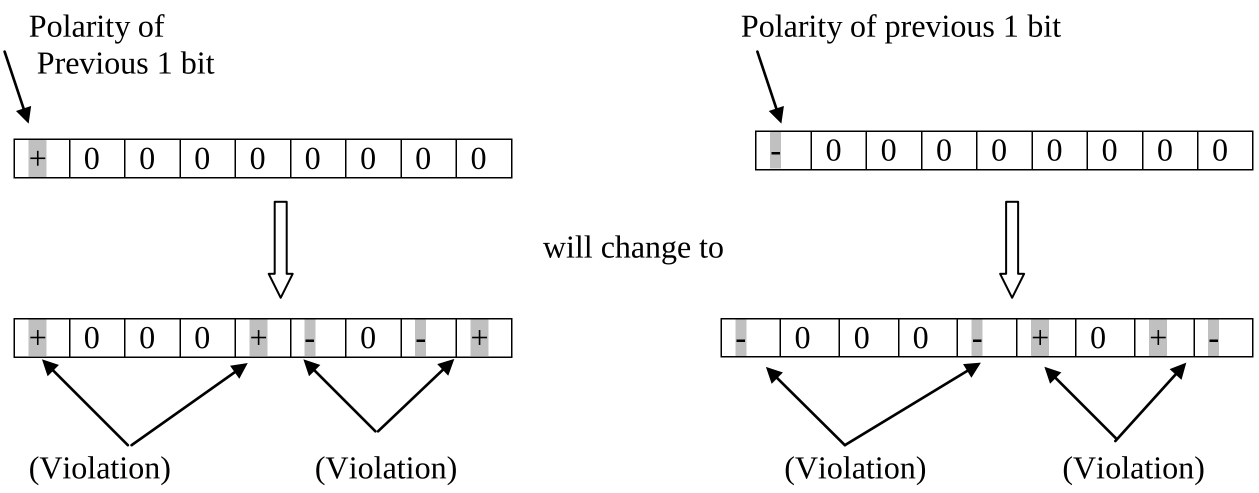
By inverting on each occurrence of a 1, bipolar AMI accomplishes two things: first, the DC component is zero, and second, a long sequence of 1s stays synchronised.

Two variations of bipolar AMI have been developed to solve the problem of synchronisation sequential 0s. The first used in North America, is called bipolar 8-zero substitution (B8ZS); the second, used in Europe and Japan, is called high-density bipolar 3 (HDB3). Both are adaptations of bipolar AMI that modify the original pattern only in the case of multiple consecutive 0s.

Bipolar 8-Zero Substitution (B8ZS)

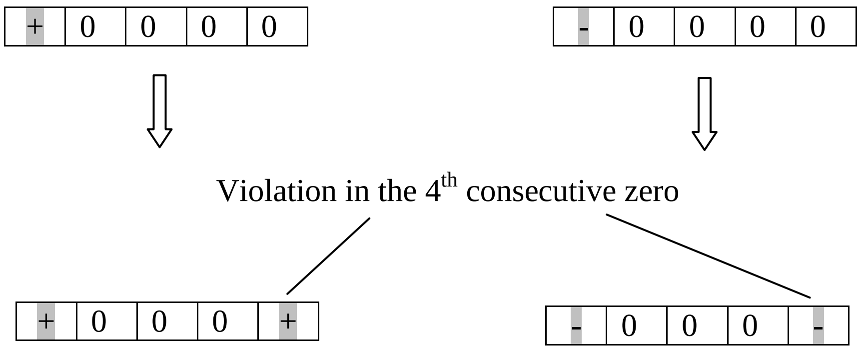
B8ZS is the convention adopted in North America to provide synchronisation of long strings of 0s. In most situations B8ZS functions identically to bipolar AMI. Bipolar AMI changes poles with every 1 it encounters. These changes provide the synchronisation needed by the receiver, but the signal does not change during a string of 0s, so synchronisation is lost. The solution provided by B8ZS is to force artificial signal changes, called violations

- In B8ZS, if eight 0s come one after another, we change the pattern in one of two ways based on the polarity of previous 1.

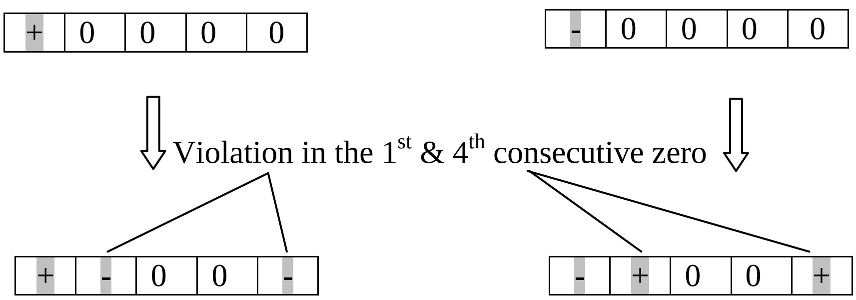


High-Density Bipolar 3 (HDB3)

In HDB3 if four 0s come one after another, we change the pattern in one of four ways based on the polarity of the previous 1 and the number of 1s since the last substitution.



If the number of 1s since the last substitution is odd



If the number of 1s since the last substitution is even

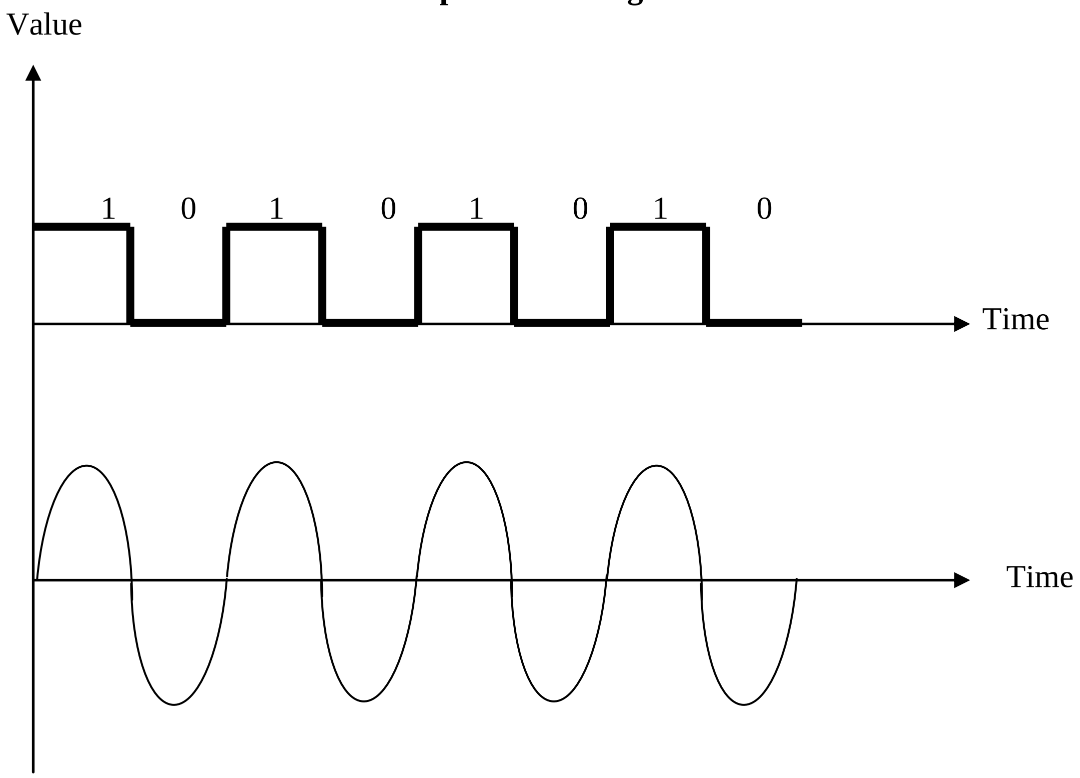
Ex

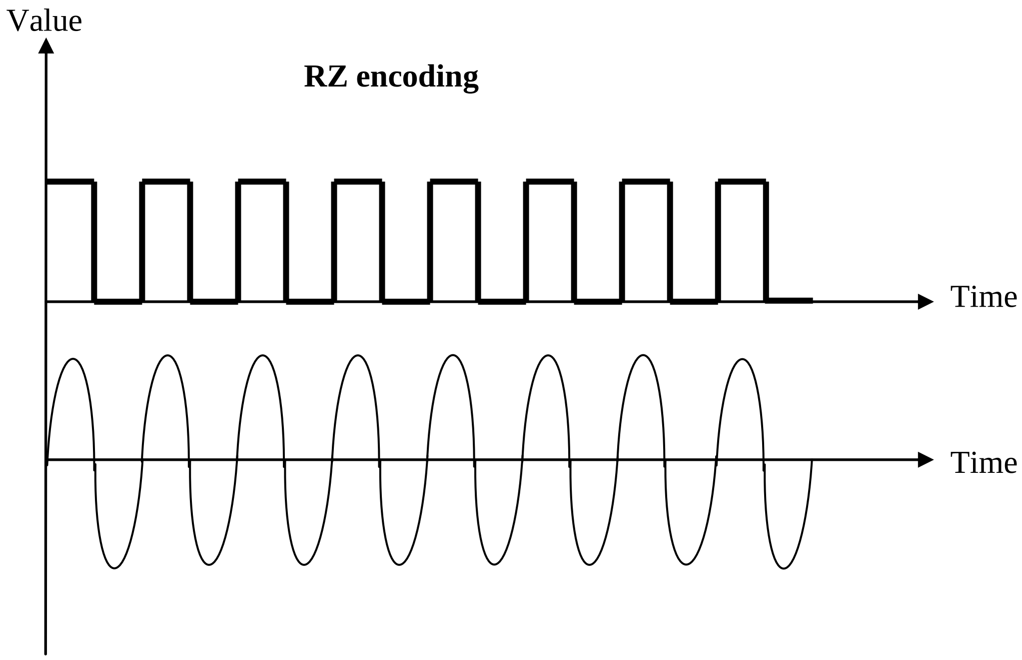
Compare the bandwidth needed for unipolar encoding and RZ encoding. Assume the worst-case scenario for both.

Solution

The worst case scenario (the situation requiring the most bandwidth) is alternating 1s and 0s for unipolar, for RZ the worst-case is all 1s.

Unipolar encoding





RZ needs twice the bandwidth of unipolar.

Ex

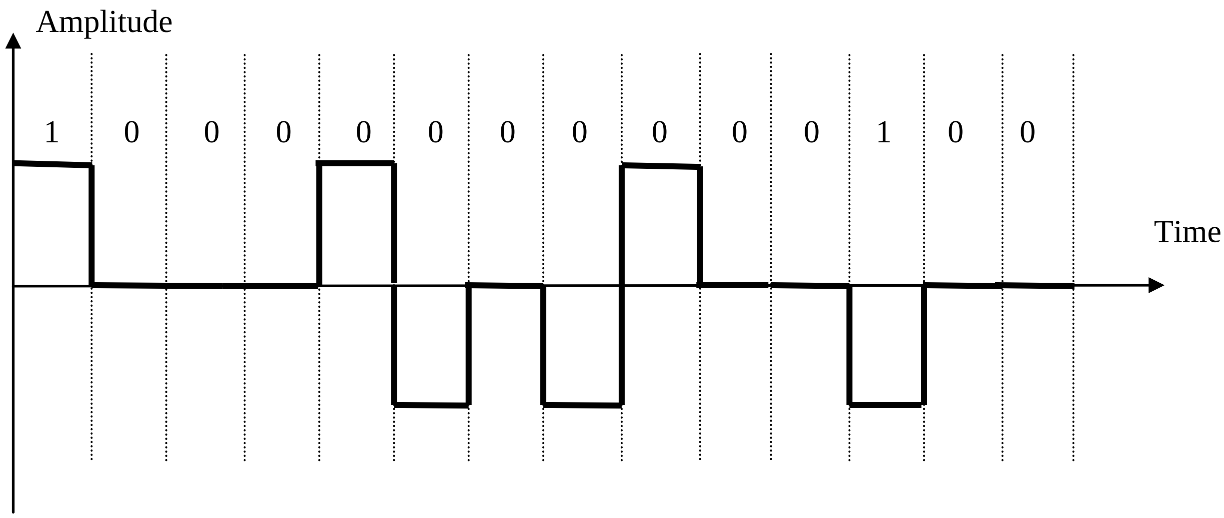
Compare the bandwidth needed for Manchester and Differential Manchester encoding. Assume the worst-case scenario for both.

Solution

The worst-case scenario for Manchester is consecutive 1s or consecutive 0s. There are two transistors for each bit (one cycle per bit). For Differential Manchester the worst – case is consecutive 0s with two transitions per each bit (one cycle per bit). The bandwidths, which are proportional to bit rate, are the same for each.

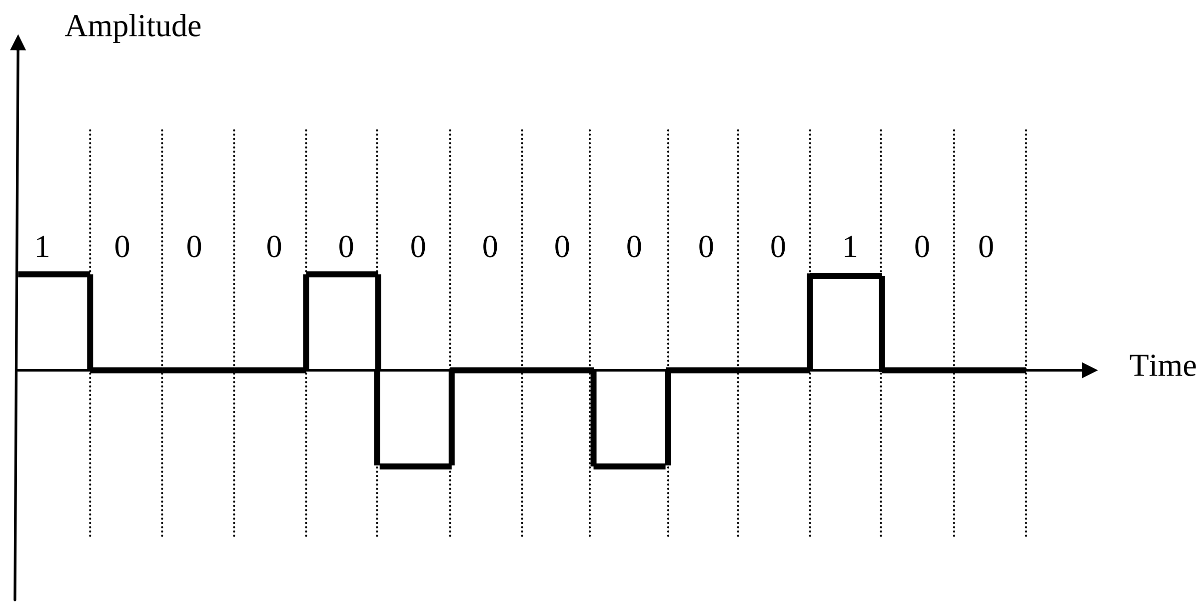
Ex

Using B8ZS, encode the bit stream 10000000000100. Assume that the polarity of the previous 1 is positive.

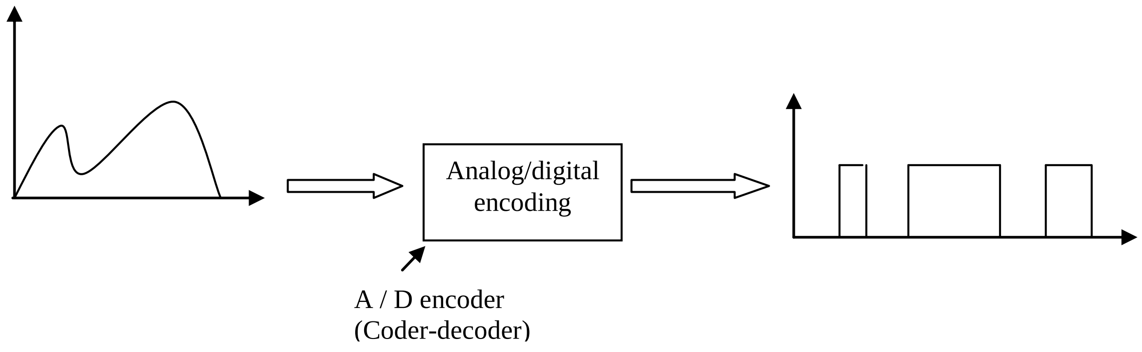


Ex

Using HDB3, encode 10000000000100. Assume that the number of 1s so far is odd and the previous 1 is positive.



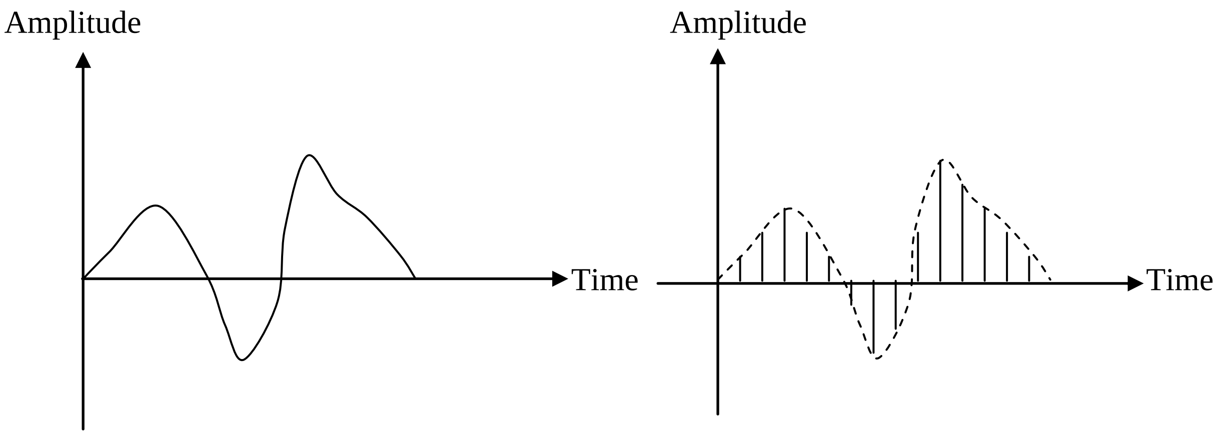
Analog-to- Digital Encoding



In analog-to-digital encoding, the information contained in a continuous wave form are represented as a series of digital pulses (1s and 0s).

Pulse Amplitude Modulation (PAM)

The first step in A/D encoding is called pulse amplitude modulation (PAM). This technique takes analog information, samples it, and generates a series of pulses based on the results of sampling. The term sampling means measuring the amplitude of the signal at equal time intervals.

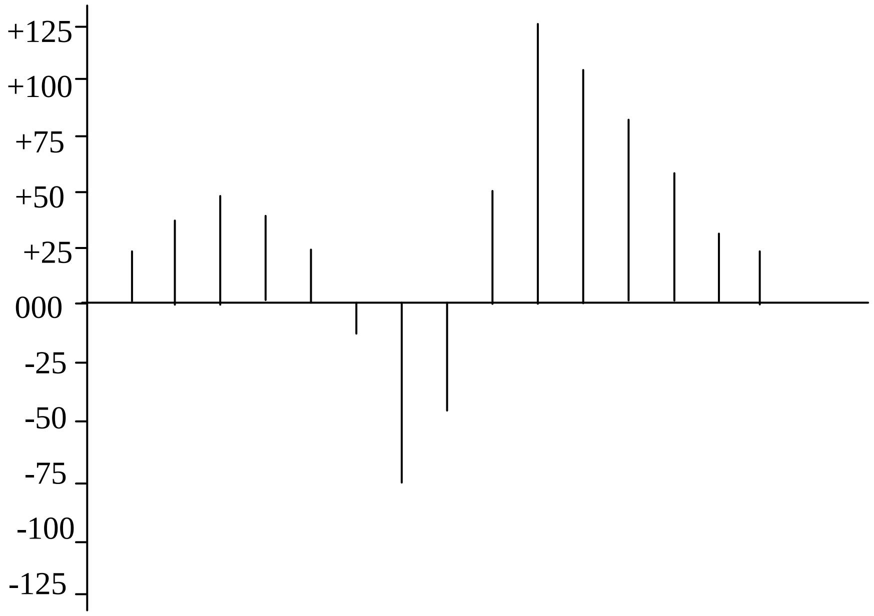


Analog Signal **PAM Signal**

In PAM, the original signal is sampled at equal intervals.
PAM has some applications, but it is not used by itself in data communications. However, it is the first step in another very popular encoding method called pulse code modulation (PCM).

Pulse Code Modulation (PCM)

PCM modifies the pulses created by PAM to create a complete digital signal. To do so, PCM first quantises the PAM pulses. Quantisation is a method of assigning integral values in a specific range to sampled instances. (The result of quantisation is presented in the following figure).

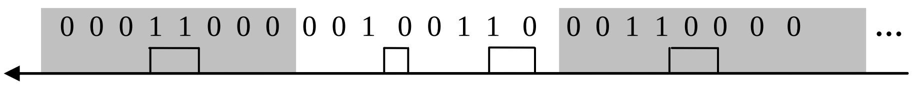


Each value is translated into its seven-bit binary equivalent. The eighth bit indicates the sign.

+24	00011000	-15	10001111	+125	01111101
+38	00100110	-80	11010000	+110	01101110
+48	00110000	-50	10110010	+90	01011010
+39	00100111	+52	00110110	+88	01011000
+26	00011010	+127	01111111	+77	01001101

The binary digits are then transformed into a digital signal using one of the digital encoding.

PCM



Direction of transfer

The result of the PCM of the original signal encoded finally into a unipolar signal.

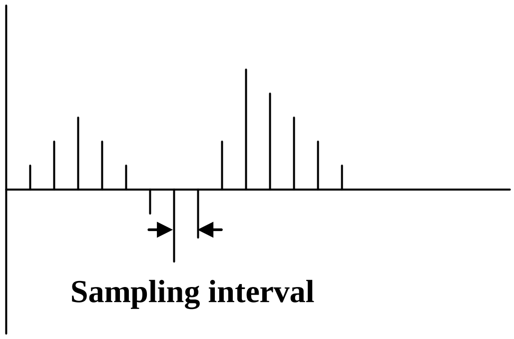
PCM is actually made up of four separate processes: PAM, quantisation, binary encoding, and digital-to-digital encoding.

PCM is the sampling method used to digitize voice in T-line transmission in the North America telecommunication system.

According to the Nyquist theorem, the sampling rate must be at least two times the highest frequency.

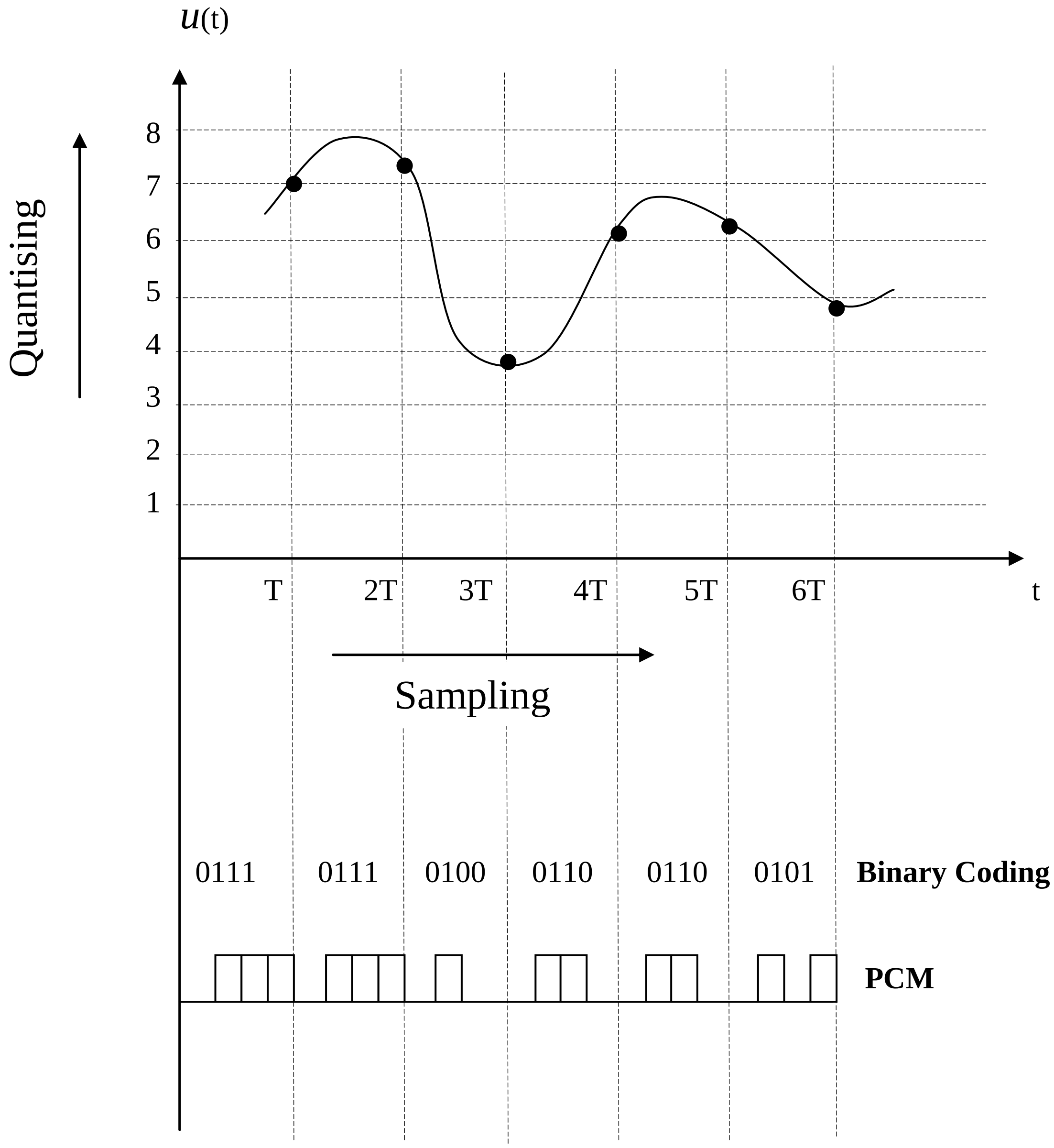
Highest frequency = x Hz

Sampling rate = $2x$ samples/second



Quantisation

Quantising is the process of rounding-off the values of the flat-top samples to certain predetermined levels.

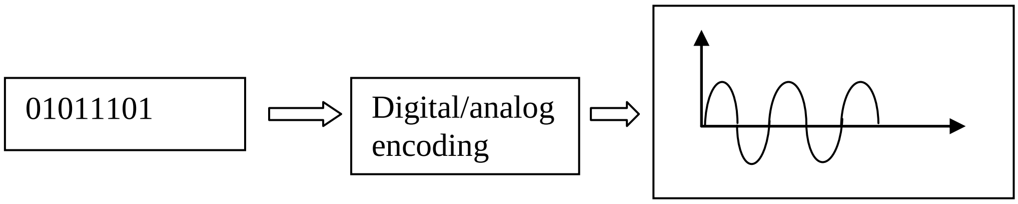


Ex What sampling rate is needed for a signal with a bandwidth of 10,000 Hz (1000 Hz to 11,000 Hz)? If the quantisation is eight bits per sample, what is the bit rate?

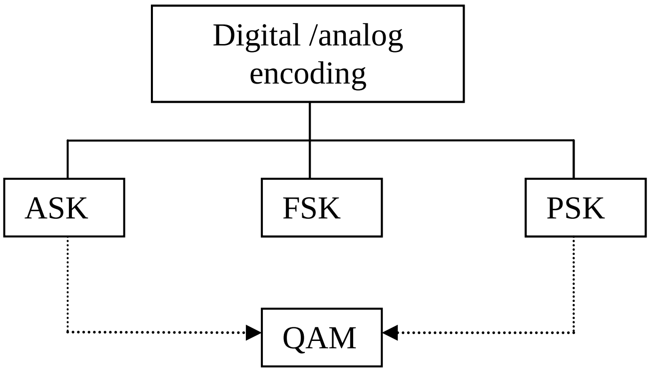
Solution

Sampling rate = 2 (11,000) = 22,000 samples/s each sample is quantised to eight bits: data rate = (22,000 samples/s) (8 bits/sample) = 176 kbps

Digital-to-Analog Encoding



Digital-to-analog encoding is the representation if digital information by an analog signal.



Quadrature Amplitude Modulation

Bit Rate and Baud Rate

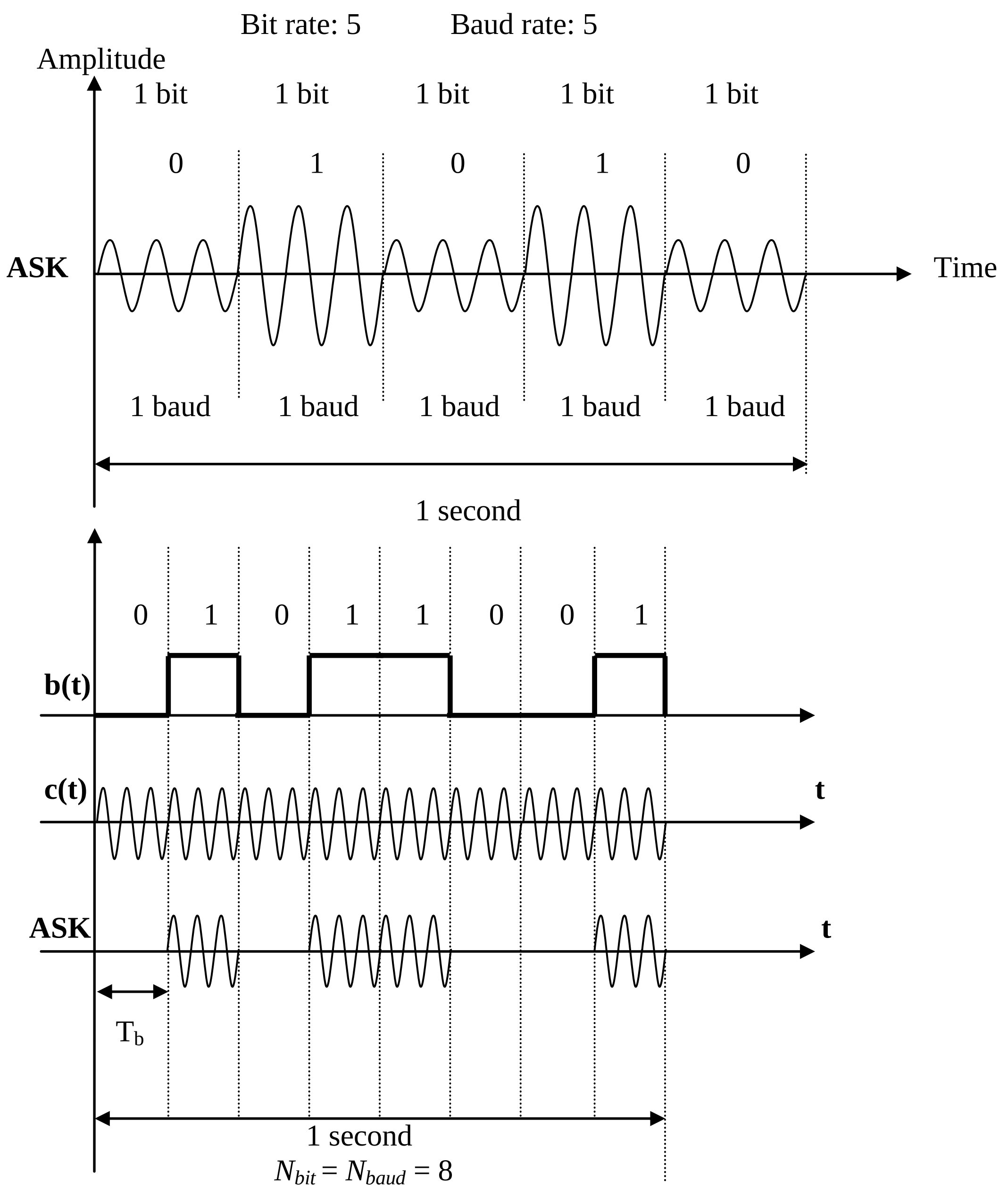
- Bit rate is the number of bits transmitted in one second.
- Baud rate refers to the number of signal units per second that are required to represent those bits.
- For computer efficiency, bit rate is more important.
- For data transmission, baud rate is more important ⇨ the fewer the signal units required, the more efficient the system, and the less bandwidth required to transmit more bits.

Carrier signal

In analog transmission the sending device produces high - frequency signal that acts as a basis for the information signal. The base signal is called the carrier signal or carrier frequency. The receiving device is tuned to the frequency of the carrier signal that it expects from the sender. Digital information is then encoded onto the carrier signal by modifying one or more of its characteristic (amplitude, frequency or phase). This kind of modification is called modulation (or shift keying) and the information signal is called a modulating signal.

Amplitude Shift Keying (ASK)

In ASK the strength of the signal is varied to represent binary 1 or 0. Both frequency and phase remain constant, while the amplitude changes.



Bit duration is the period of time that defines one bit. The peak amplitude of the signal, during each bit duration, is constant and its value depends on the bit (0 or 1). The transmission speed using ASK is limited by the physical characteristics of the transmission medium.

Bandwidth for ASK

$$BW = (1 + d)*N_{baud}$$

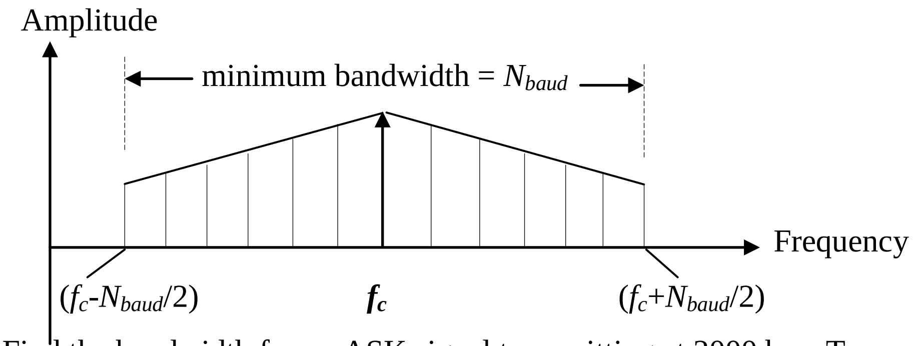
Where

BW is the bandwidth

N_{baud} is the baud rate

d is a factor related to the condition of the line (with min. value of 0)

- The minimum bandwidth required for transmission is equal to the baud rate.



Ex Find the bandwidth for an ASK signal transmitting at 2000 bps. Transmission is in half-duplex mode.

Solution

In ASK baud rate = bit rate

$N_{baud} = 2{,}000$

An ASK signal requires a bandwidth equal to its baud rate:

$\Rightarrow$ BW = 2,000 Hz.

Ex Given a bandwidth of 10,000 Hz (1,000 to 11,000 Hz), draw the full-duplex ASK diagram of the system. Find the carriers and the bandwidth in each direction. Assume there is no gap between the bands in two directions.

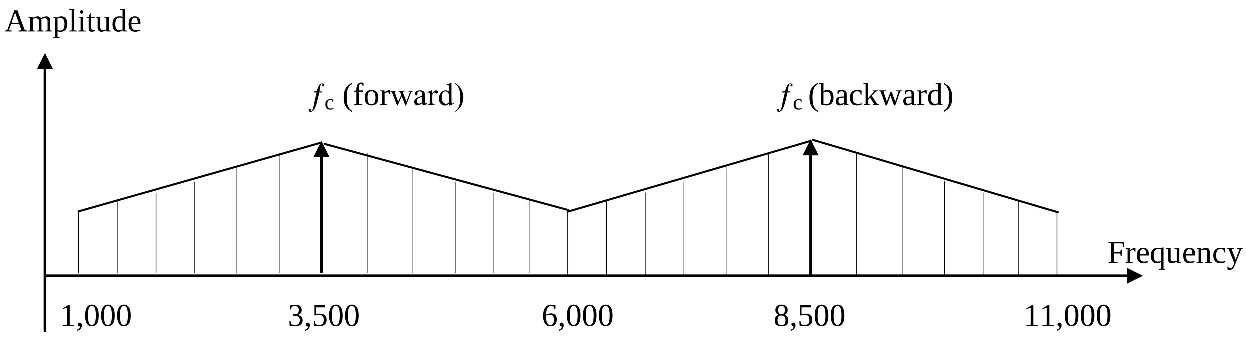
Solution

For full-duplex ASK the bandwidth for each direction is BW=10,000/2=5000 Hz.

The carrier frequencies can be chosen at the middle of each band

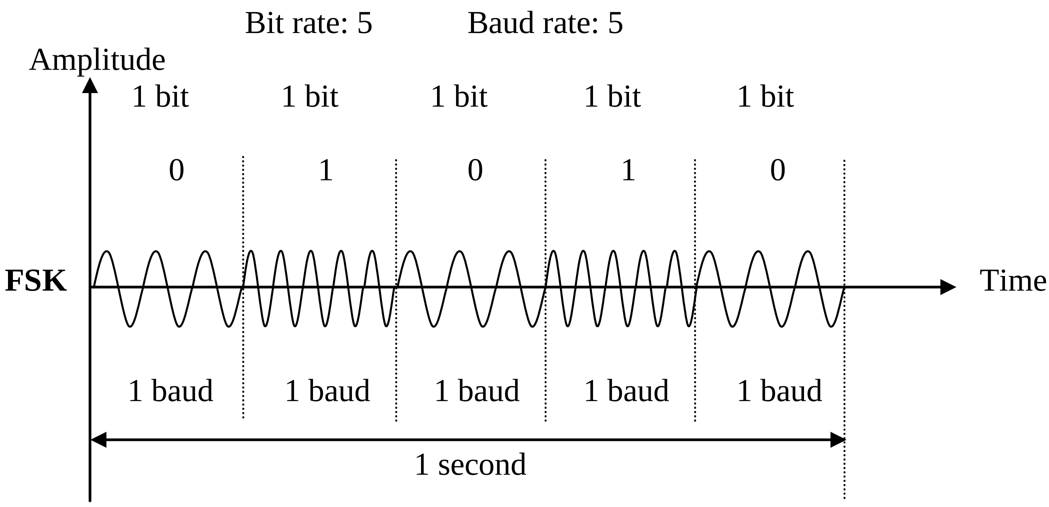
$f_{c\ (forward)} = 1{,}000 + 5{,}000/2 = 3{,}500$ Hz

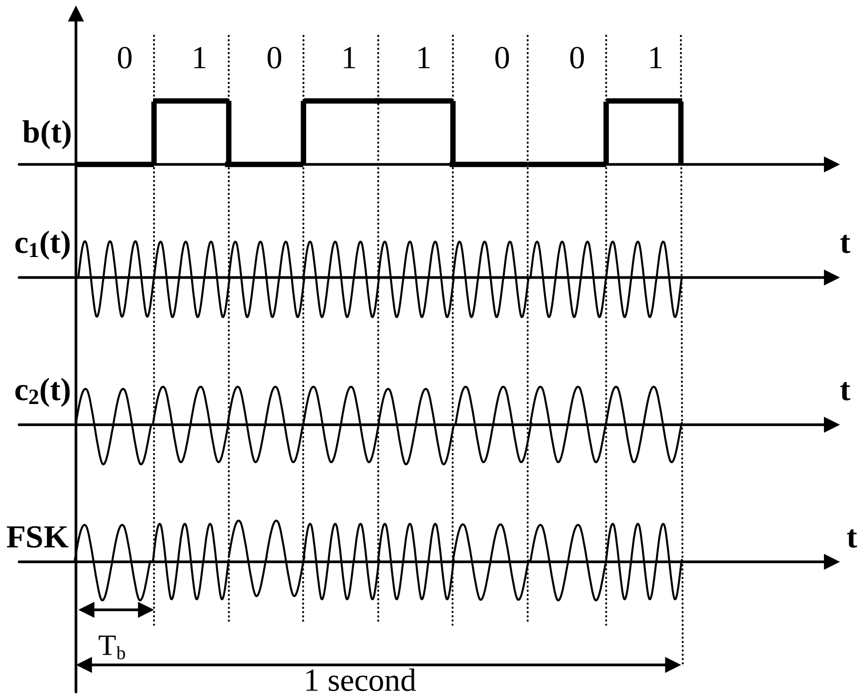
$f_{c\ (backward)} = 11{,}000 - 5{,}000/2 = 8{,}500$ Hz



Frequency Shift Keying (FSK)

In frequency shift keying (FSK), the frequency of the signal is varied to represent binary 1 or 0. The frequency of the signal during each bit duration is constant and its value depends on the bit (0 or 1): both peak amplitude and phase remain constant.



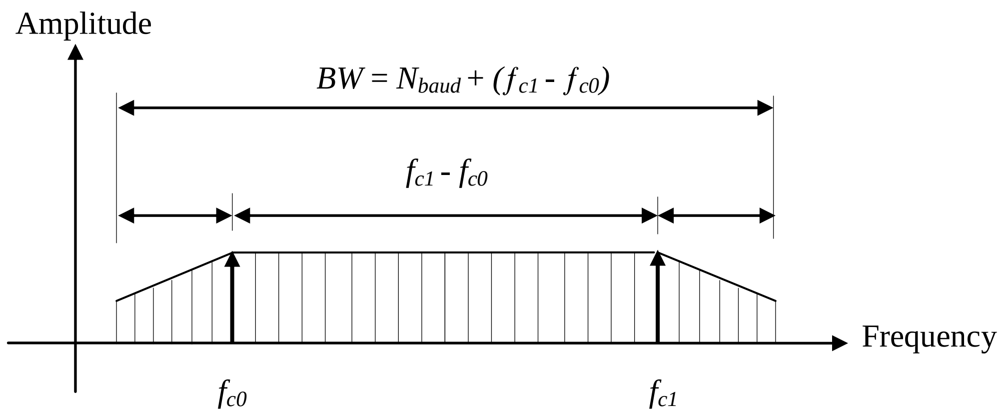


$N_{bit} = N_{baud} = 8$

FSK avoids most of the noise problems of ASK. The limiting factors of FSK are the physical capabilities of the carrier.

Bandwidth for FSK

FSK spectrum can be considered as the combinations of two ASK spectra centred on f_{c0} and f_{c1}. The bandwidth required for FSK transmission is equal to the baud rate of the signal plus the frequency shift (difference between the two carrier frequencies).



$$\boxed{BW = N_{baud} + (f_{c1} - f_{c0})}$$

Ex Find the bandwidth for an FSK signal transmitting at 2,000 bps. Transmission is in half-duplex mode and the carriers must be separated by 3,000 Hz.

Solution

$$BW = N_{baud} + (f_{c1} - f_{c0})$$
$$= 2{,}000 + 3{,}000 = 5{,}000 \text{ Hz.}$$

Ex Find the maximum bit rate for an FSK signal if the bandwidth of the medium is 12,000 Hz and the distance between the two carriers must be at least 2,000 Hz. Transmission is in full-duplex mode.

Solution

Because the transmission is in full-duplex, only 6,000 Hz is allocated for each direction, for FSK, if f_{c1} and f_{c0} are the carrier frequencies,

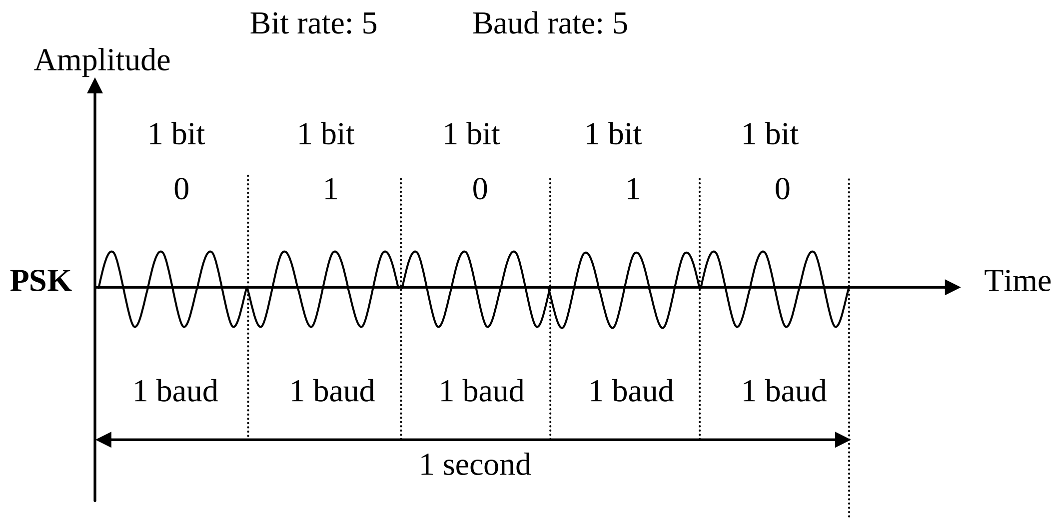
$$BW = N_{baud} + (f_{c1} - f_{c0})$$
$$N_{baud} = BW - (f_{c1} - f_{c0})$$

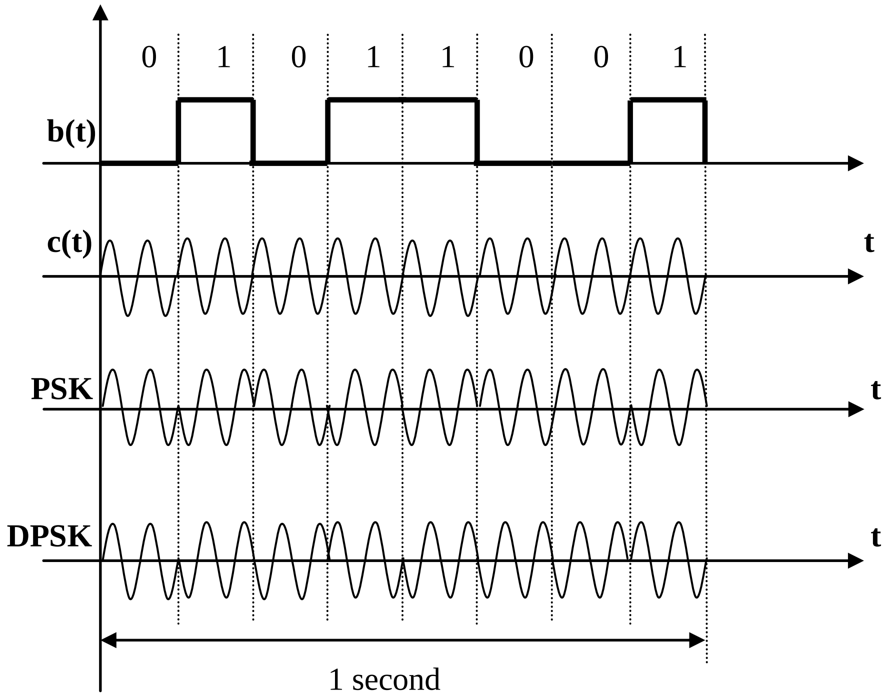
= 6,000 – 2,000 = 4,000

But because the baud rate is the same is bit rate, the bit rate is 4,000 bps.

Phase Shift Keying (PSK)

In the PSK, the phase is varied to represent binary 1 or 0. Both peak amplitude and frequency remain constant as the phase changes. The phase of the signal during each bit duration, is constant and its value depends on the bit (0 or 1).





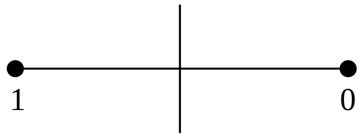
$N_{bit} = N_{baud} = 8$

DPSK eliminates the need for a coherent reference signal at the receiver by combing two basic operations at the transmitter:

- Differential encoding of the input data
- PSK ⟹ DPSK
- To send symbol 1 we phase advance the current signal waveform by 180°.
- To send symbol 0 we leave the phase of the current signal waveform unchanged.

PSK Constellation

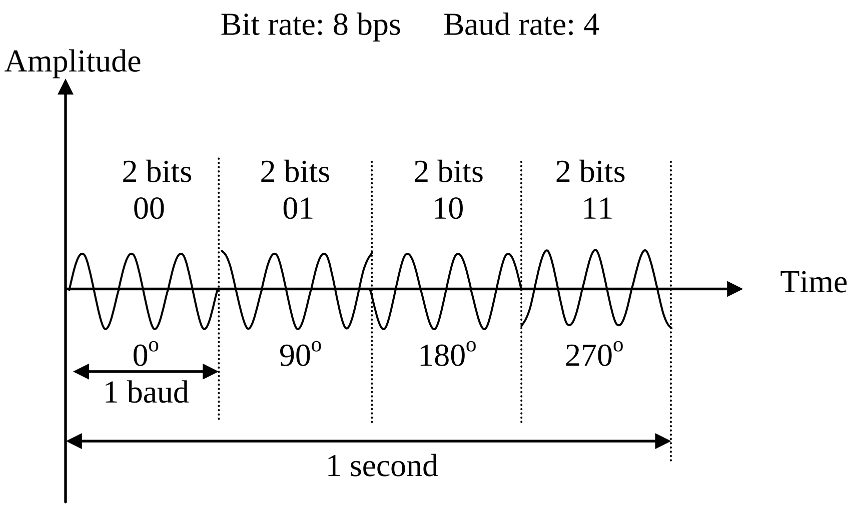
Bit	Phase
0	0^o
1	180°



Constellation diagram

The above method is often called 2-PSK, or binary PSK, because two different phases (0° and 180°) are used in the encoding.

PSK is not susceptible (easily influenced) to the noise degradation that affects ASK, nor to the bandwidth limitations of FSK. This means that smaller variations in the signal can be detected reliably by the receiver. Therefore instead of utilising only two variations of a signal, each representing one bit, we can use four variations and let each phase shift represent two bits.



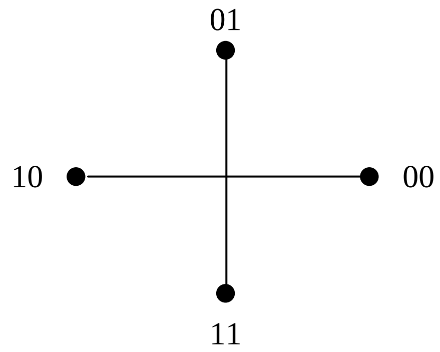
4-PSK (Quadrature-PSK)

This technique is called 4-PSK or Q-PSK. The pair of bits represented by each phase is called a dibit.

Data can be transmitted two times as fast using 4-PSK as using 2-PSK.

4–PSK characteristics

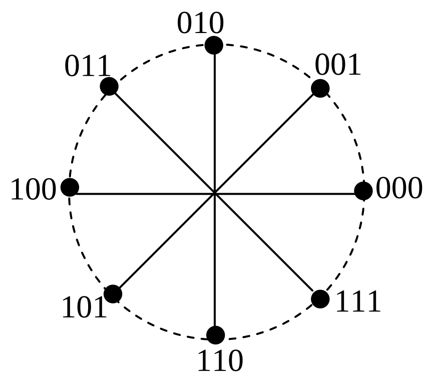
Dibit	Phase
00	0^o
01	90^o
10	180^o
11	270^o



Constellation diagram

8–PSK Characteristics

Tribit	Phase
000	0^o
001	45^o
010	90^o
011	135^o
100	180^o
101	225^o
110	270^o
111	315^o



Constellation diagram

Bit rate of 8-PSk is three as that of 2-PSK

Bandwidth for PSK

The min. BW required for PSK transmission is the same as that required for ASK transmission.

$$\text{BW} = N_{baud}$$

***Ex*:** Find the bandwidth for a 4-PSK signal transmitting at 2,000 bps. Transmission is in half-duplex mode.

Solution

For 4-PSK the baud rate is half of the bit rate.

The baud rate is therefore 1,000. A PSK signal requires a bandwidth equal to its baud rate. Therefore the bandwidth is 1,000 Hz.

***Ex*:** Given a bandwidth of 5,000 Hz for an 8-PSK signal, what are the baud rate and bit rate?

Solution

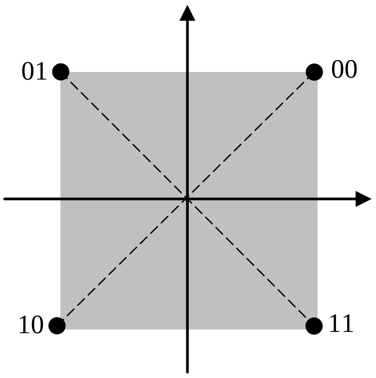
For PSK the baud rate is the same as the bandwidth, which means the baud rate is 5,000. But in 8-PSK, the bit rate is three times the baud rate. So the bit rate is 15,000 bps.

Quadrature Amplitude Modulation (QAM)

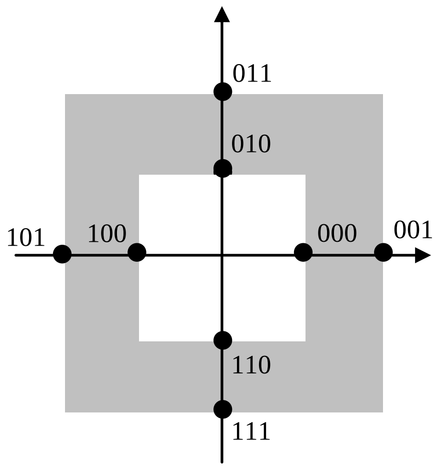
QAM means combining ASK and PSK in such a way that we have maximum contrast between each bit, dibit, tribit, quadbit, and so on.

Possible variation of QAM is numerous; theoretically any measurable number of changes in amplitude can be combined with any measurable number of changes in phase.

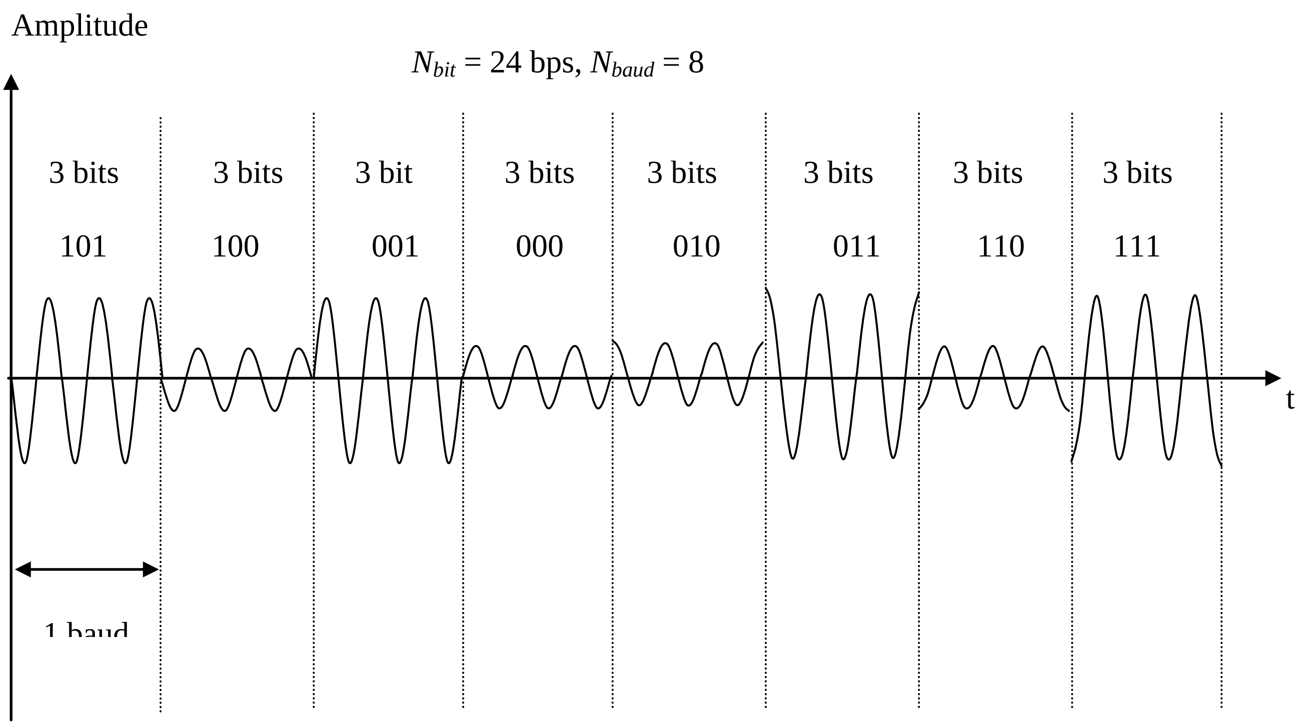
The figure below shows the constellation diagrams of 4-QAM and 8-QAM. In both cases the number of amplitude shifts is less than the number of phase shifts. Because amplitude changes are susceptible to noise and require greater shift differences than do phase changes, the number of phase shifts used by a QAM system is always larger than the number of amplitude shifts.



4-QAM
Constellation Diagram



8-QAM
Constellation Diagram

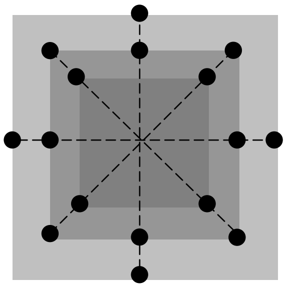


The output signal of the 8-QAM modem for data

b (t) = 101 100 001 000 010 011 110 111

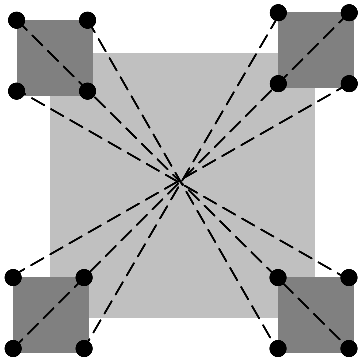
Three popular 16-QAM configurations are shown bellow:

4 amplitudes, 8 phase



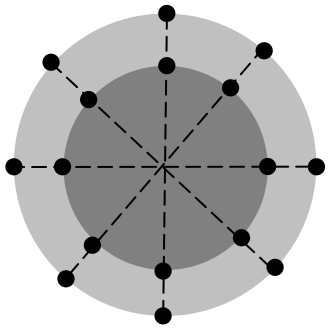
16-QAM

3 amplitudes, 12 phases



16-QAM

2 amplitude, 8 phase



16-QAM

Since amplitude shift is more susceptible to noise, the greater the ratio of phase shifts to amplitude, the greater the immunity to noise.

The second example, three amplitudes and 12 phases, handles noise best. The first example, 4 amplitudes and 8 phases, is the OSI (Open Systems Interconnection) recommendation. Several QAM designs link specific amplitudes with specific phases. This means that even with noise problems associated with amplitude shifting, the meaning of a shift can be recovered from phase information.

Bandwidth for QAM

The minimum bandwidth required for QAM transmission is the same as that required for ASK and PSK transmission. QAM has the same advantages of PSK over ASK.

Bit/Baud Comparison

Assuming that an FSK signal over voice-grade phone lines can send 1200 bps, the bit rate is 1200 bps. Each frequency shift represents a single bit; so it requires 1200 signal elements to send 1200 bits. Its baud rate, therefore, is also 1200. Each signal variation in 8-QAM system, however, represents three bits. So a bit rate of 1200 bps, using 8-QAM, has a baud rate of only 400.

As the figure below shows, a dibit system has a baud rate of one-half the bit rate, a tribit system has a baud rate of one-third the bit rate, a quadbit system has a baud rate of one-fourth the bit rate.

Bit

N_{baud}= N N_{bit}= N

0	0	1	0	1	0	0	0	1	0	1	0	1	1	1	0

Dibit

N_{baud}= N N_{bit}= 2N

0 0	1 0	1 0	0 0	1 0	1 0	1 1	1 0

Tribit

N_{baud}= N N_{bit}= 3N

0 0 1	0 1 0	0 0 1	0 1 0	1 1 1	0

Quadbit

N_{baud}= N N_{bit}= 4N

0 0 1 0	1 0 0 0	1 0 1 0	1 1 1 0

Bit/ Baud Rate Comparison

Modulation	*Units*	*Bits/Bauds*	*Baud Rate*	*Bit Rate*
ASK, FSK, 2-PSK	**Bit**	**1**	**N**	**N**
4-PSK, 4-QAM	**Dibit**	**2**	**N**	**2N**
8-PSK, 8-QAM	**Tribit**	**3**	**N**	**3N**
16-QAM	**Quadbit**	**4**	**N**	**4N**
32-QAM	**Pentabit**	**5**	**N**	**5N**
64-QAM	**Hexabit**	**6**	**N**	**6N**
128-QAM	**Septabit**	**7**	**N**	**7N**
256-QAM	**Octabit**	**8**	**N**	**8N**

Ex

A constellation diagram consists of eight equally spaced points on a circle. If the bit rate is 4800 bps, what is the baud rate?

Solution

The constellation indicates 8-PSK with points 45^o apart. Since $2^3 = 8$, three bits are transmitted with each signal element. Therefore, the baud rate is

$$4800/3 = 1600 \text{ baud}$$

Ex

Compute the baud rate for a 72,000 bps 64-QAM.

Solution

A 64-QAM signal means that there are six bits per signal elements since $2^6 = 64$. Thus,

$$72{,}000/6 = 12{,}000 \text{ baud}$$

Ex

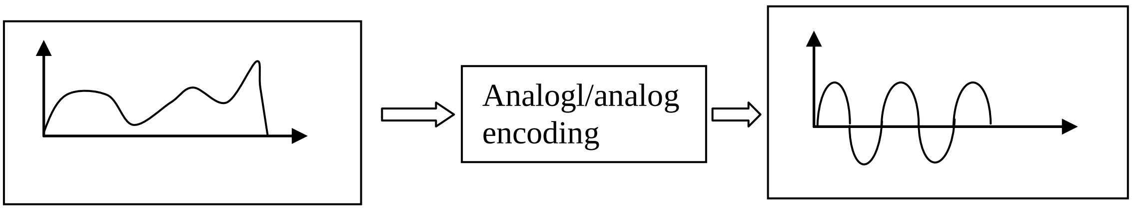
Compute the bit rate for a 1,000 baud 16-QAM signal.

Solution

A 16-QAM signal means that there are four bits per signal elements since $2^4 = 16$. Thus,

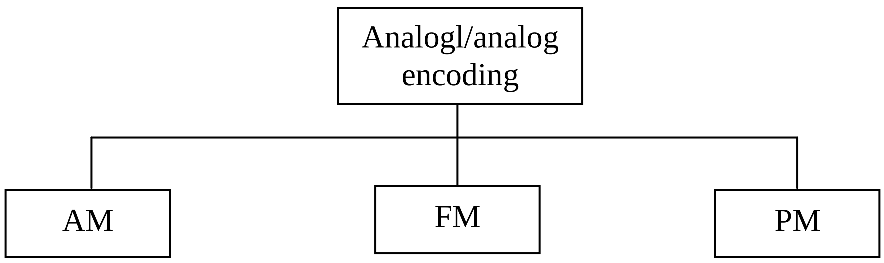
$$(1{,}000)(4) = 4{,}000 \text{ bps}$$

Analog-to-Analog-Encoding



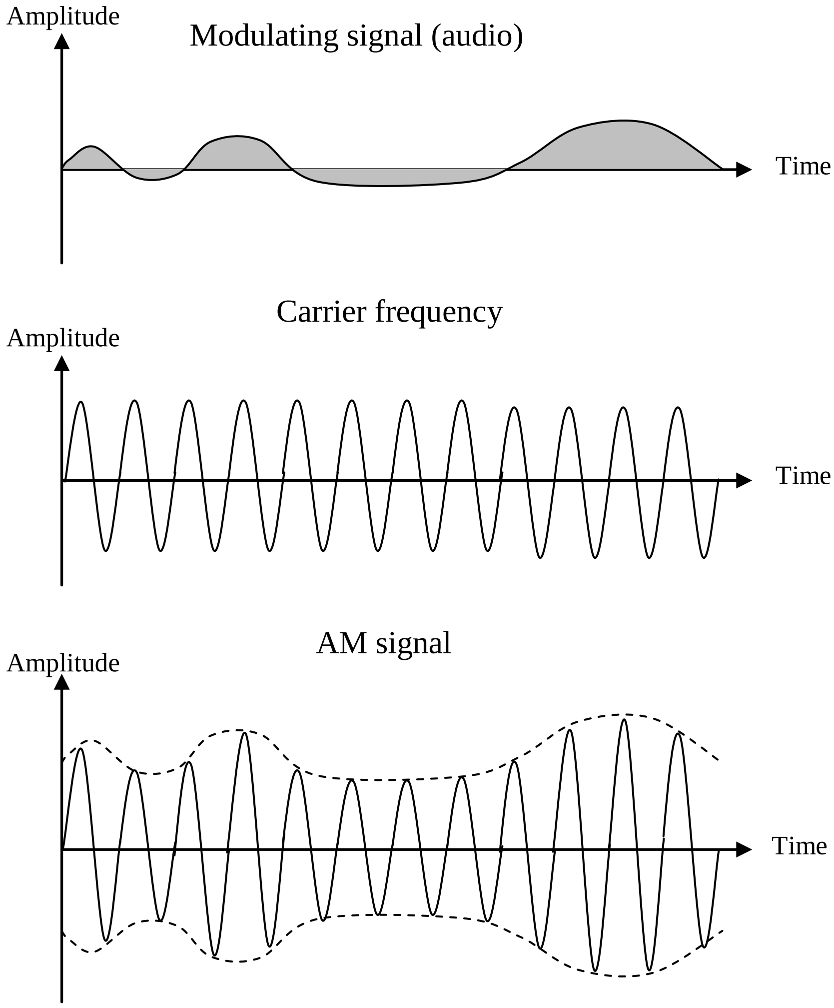
Analog-to-analog encoding is the representation of analog information by an analog signal. (eg. Radio communication).

Analog-to-analog modulation can be accomplished in three ways: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM)



Amplitude Modulation

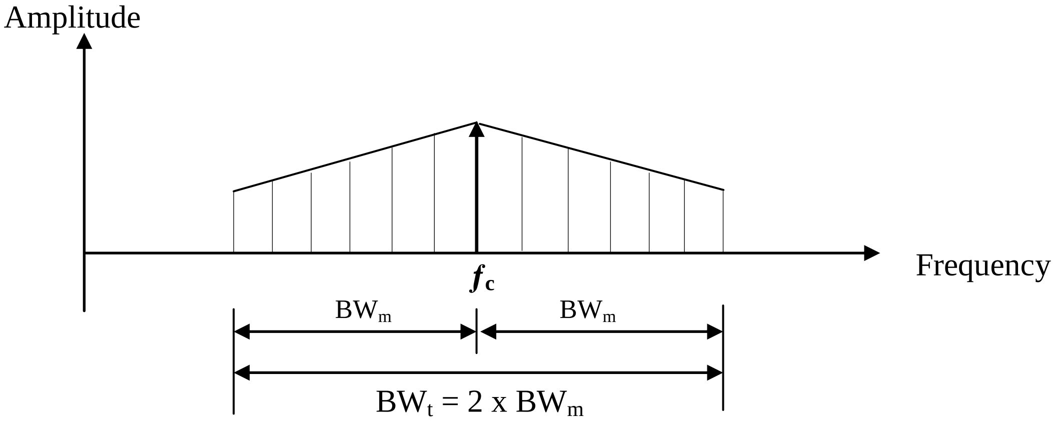
In AM transmission, the carrier signal is modulated so that its amplitude varies with the changing amplitudes of the modulating signal. The frequency and phase of the carrier remain the same; only the amplitude changes to follow variations in the information. The modulating signal becomes the envelope of the carrier.



AM Bandwidth

The bandwidth of an AM signal is equal to twice the bandwidth of the modulating signal and covers a range centred on the carrier frequency. The total bandwidth required for AM can be determined from the bandwidth of the audio signal:

$$\mathbf{BW_t = 2 \times BW_m}$$



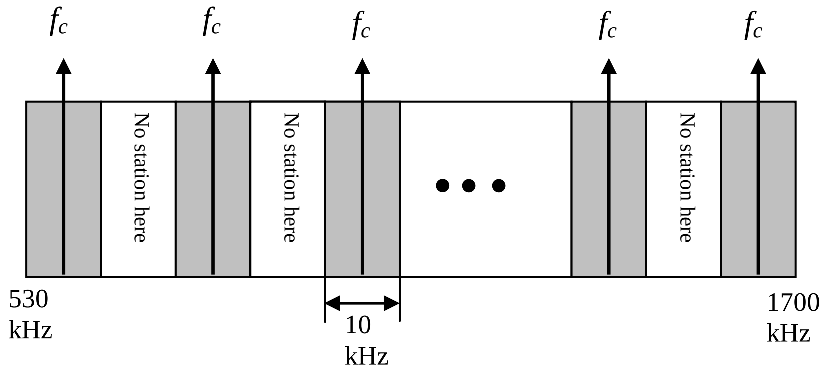
BW_m = Bandwidth of the modulating signal (audio)

BW_t = Total bandwidth (radio)

f_c = Frequency of the carrier

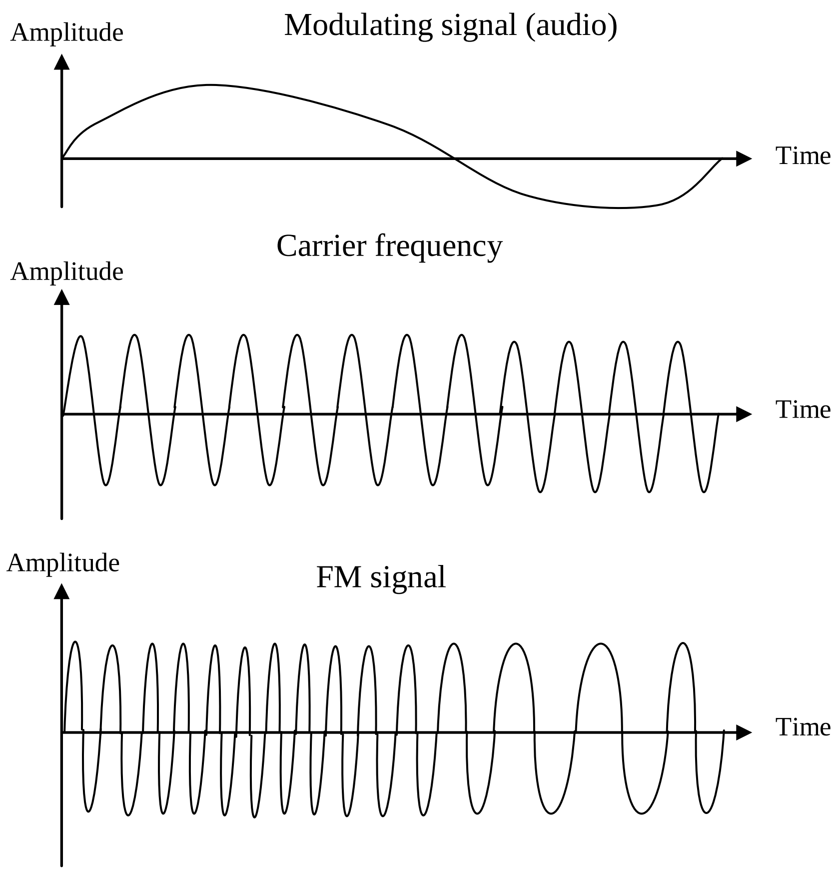
The bandwidth of an audio signal (speech & music) is usually 5 kHz. Therefore, an AM radio station needs a minimum bandwidth of 10 kHz. In fact, the Federal Communications Commission (FCC) allows 10 kHz for each AM station.

AM stations are allowed carrier frequencies anywhere between 530 and 1700 kHz (1.7 MHz). However, each station's carrier frequency must be separated from those on either side by at least 10 kHz (one AM bandwidth) to avoid interference.



Frequency Modulation (FM)

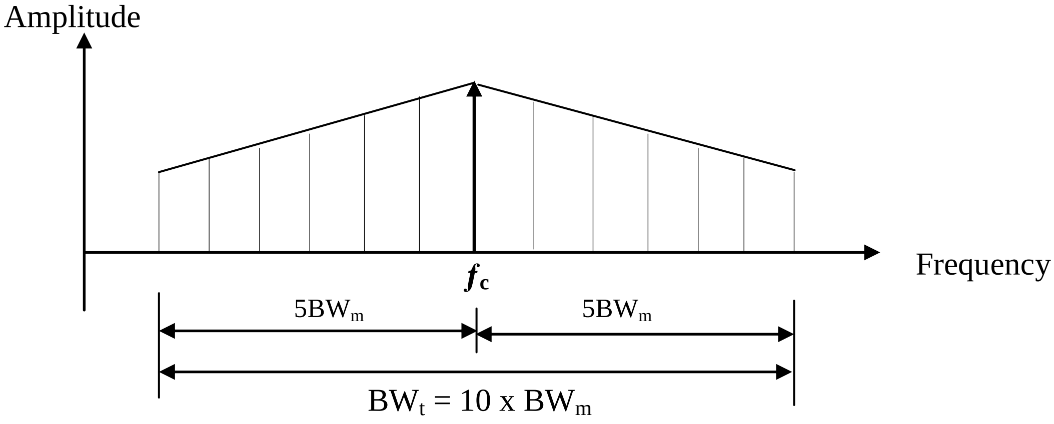
In FM transmission, the frequency of the carrier signal is modulated to follow the changing voltage level (amplitude) of the modulating signal. The peak amplitude and phase of the carrier signal remain constant, but as the amplitude of the information signal changes, the frequency of the carrier changes correspondingly.



FM Bandwidth

The bandwidth of an FM signal is equal to 10 times the bandwidth of the modulating signal and, like AM bandwidth, covers a range centred on the carrier frequency. The total bandwidth required for FM can be determined from the bandwidth of the audio signal:

$$\mathbf{BW_t = 10 \times BW_m}$$



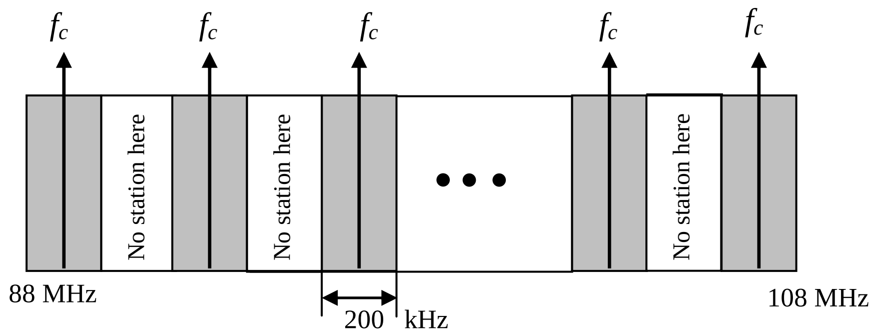
BW_m = Bandwidth of the modulating signal (audio)

BW_t = Total bandwidth (radio)

f_c = Frequency of the carrier

The bandwidth of an audio signal (speech & music) broadcast in stereo is almost 15 kHz. Therefore, each FM radio station needs a minimum bandwidth of 150 kHz. The FCC allows 200 kHz (0.2 MHz) for each FM station to provide some room for guard bands.

FM stations are allowed carrier frequencies anywhere between 88 and 108 MHz. However, stations must be separated from by at least 200 kHz to avoid overlapping.



Phase Modulation (PM)

Due to simpler hardware requirements, PM is used in some systems as an alternative to FM. In PM transmission, the phase of the carrier signal is modulated to follow the changing voltage level of the modulating signal. The peak amplitude and frequency of the carrier signal remain constant, but as the amplitude of the information signal changes, the phase of carrier changes correspondingly. The analysis and final result (modulating signal) are similar to those of FM.

TRANSMISSION CODES

Transmission Codes

Binary-Coded Decimal (also called 8421 BCD)

In BCD, four bits are used to encode one decimal character. Four bits give 16 binary combinations. Since there are 10 decimal characters, 0 through 9, only 10 of the 16 possible combinations are necessary for encoding in BCD. The remaining 6 combinations are said to be invalid.

Decimal	BCD
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

$$\left.\begin{array}{l}1010\\1011\\1100\\1101\\1110\\1111\end{array}\right\}\text{Not valid in BCD}$$

Ex

Convert 367_{10} to BCD

Solution

$$367_{10} = 0011\ \ 0110\ \ 0111$$

Ex

Convert 1249_{10} to BCD

Solution

$$1249_{10} = 0001\ \ 0010\ \ 0100\ \ 1001$$

Ex

Convert 58_{10} to BCD

Solution

$$58_{10} = 0101\ \ 1000$$

BCD Addition:

Straight binary addition is performed as long as the result does not exceed a decimal value of 9.

Ex

Add the decimal numbers 3 and 4 in BCD

Solution

$$\begin{array}{r} 3 \\ +\,4 \\ \hline 7 \end{array} \qquad \begin{array}{r} 0011 \\ 0100 \\ \hline 0111 \end{array}$$

Ex

Add the decimal numbers 63 and 24 in BCD

Solution

$$\begin{array}{r} 63 \\ +\,24 \\ \hline 87 \end{array} \qquad \begin{array}{r} 0110\ \ 0011 \\ 0010\ \ 0100 \\ \hline 1000\ \ 0111 \end{array}$$

When the sum of two numbers exceeds 9, an invalid BCD number is obtained. The invalid number can be converted to a valid number by adding 0110 (6) to it.

Ex

Add the decimal numbers 9 and 6 in BCD

Solution

$$\begin{array}{r} 9 \\ +\,6 \\ \hline 15 \end{array} \qquad \begin{array}{r} 1001 \\ 0110 \\ \hline 1111 \end{array}$$

1111…not valid in BCD
+ 0110…add 6 for correction
0001 0101…correct BCD number

Ex

Add the decimal numbers 46 and 79 in BCD

Solution

$$\begin{array}{r} 46 \\ +\,79 \\ \hline 125 \end{array} \qquad \begin{array}{rrr} & 0100 & 0110 \\ & 0111 & 1001 \\ \hline & 1011 & 1111 \\ + & 0110 & 0110 \\ \hline 0001 & 0010 & 0101 \\ 1 & 2 & 5 \end{array}$$

Excess-3 Code

Excess-3 code is very similar to 8421 BCD code. The only difference is that 3 is added to the decimal before it is encoded into a four-word.

Decimal	BCD	Excess-3
0	0000	0011
1	0001	0100
2	0010	0101
3	0011	0110
4	0100	0111
5	0101	1000
6	0110	1001
7	0111	1010
8	1000	1011
9	1001	1100

0000
0001
0010
1101
1110
1111

Not valid in Excess-3

Ex

Add the decimal numbers 9 and 7 in Excess-3

Solution

$$\begin{array}{rr} 9 & 1100 \\ +\ 7 & 1010 \\ \hline 16 & 1\ 0110 \end{array}$$

Gray Code

The disadvantage of the previous codes is that several bits change state between adjacent counts. The Gray code is unique in that successive counts result in only one bit change. For example, 7 (0111) to 8 (1000) in binary, or BCD, all four bits change state. In Gray code, however, 7 (0100) to 8 (1100) require a single bit change.

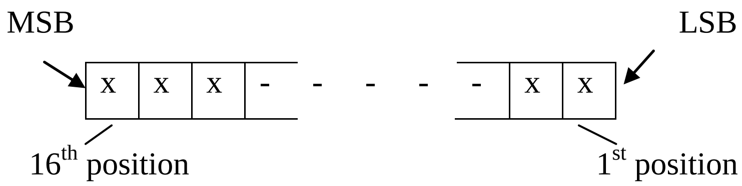
The switching noise generated by the associated circuits may be intolerable in some environments. The same change with Gray code undergoes only a single bit change consequently, less noise is generated. Shaft encoders used for receiver tuning dials often use Gray code.

The Gray code is widely used for encoding the position of the rotary shaft and for data transmission using PSK.

Decimal	Binary	Gray Code
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111

Binary-to-Gray Conversion

- The given binary code is shifted to the right by one bit.
- Discard the last bit (the LSB) from the obtained bits.
- Exclusive-ORing the given and obtained bits result in the equivalent Gray code.



Ex

Compute the Gray code for the binary number 11010

Solution

Binary code		1 1 0 1 0
	$\oplus$	1 1 0 1
Gray code		**1 0 1 1 1**

Ex

Compute the Gray code for the binary number 10001101

Solution

Binary code		1 0 0 0 1 1 0 1
	$\oplus$	1 0 0 0 1 1 0
Gray code		**1 1 0 0 1 0 1 1**

Gray-to-Binary Conversion

- The first bit, the leftmost of the given Gray code, becomes the MSB of the Binary code.
- Exclusive-ORing the second Gray code bit with the MSB of the binary code yields the second binary bit.
- Exclusive-ORing the third Gray code bit with the second binary code yields the third binary bit.
- Exclusive-ORing the fourth Gray code bit with the third binary code yields the fourth binary bit. And so on.

Ex

Compute the binary code for the Gray code 101101

Solution

Gray code	1	0	1	1	0	1
Binary code	**1**	**1**	**0**	**1**	**1**	**0**

Ex

Compute the binary code for the Gray code 11001011

Solution

Gray code	1	1	0	0	1	0	1	1
Binary code	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**1**

Binary Numbers

The binary numbering system provides the basis for all computer operations. Computers work by manipulating electrical current on and off. The binary system uses two symbols, 0 and 1. Also called base 2.

Binary weights

Position	**Fifth**	**Fourth**	**Third**	**Second**	**First**
Weight	2^4 (16)	2^3 (8)	2^2 (4)	2^1 (2)	2^0 (1)

Ex

1	1	0	1	digits
8	4	2	1	weights
8	4	0	1	results

+

13

Octal Numbers

The octal numbering system is used by computer programmers to represent binary numbers in compact form. Also called base 8.

Octal numbers use 8 symbols: 0,1,2,3,4,5,6,7.

Octal weights

Position	**Fifth**	**Fourth**	**Third**	**Second**	**First**
Weight	8^4 (4096)	8^3 (512)	8^2 (64)	8^1 (8)	8^0 (1)

Ex

3	4	7	1	digits
512	64	8	1	weights
1,536	256	56	1	results

+

1,849

Hexadecimal Numbers

Hexadecimal numbering system, like octal, is used by computer programmers to represent binary numbers in compact form. Also called base 16.

Hexadecimal uses 16 symbols: 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F.

Hexadecimal weights

Position	**Fifth**	**Fourth**	**Third**	**Second**	**First**
Weight	16^4 (65,536)	16^3 (4,096)	16^2 (256)	16^1 (16)	16^0 (1)

Ex

3	4	7	1	digits
4,096	256	16	1	weights
12,288	1,024	112	1	results

+

13,425

Decimal	Binary	Octal	Hexadecimal
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

Transformations

- From Other Systems to Decimal

a) From Binary to Decimal

1	0	0	1	1	1	0	Binary
64	32	16	8	4	2	1	weights
64	0	0	8	4	2	0	weighted results

+

78

b) Hexadecimal to Decimal

4	E	Hexadecimal
16	1	weights
64	14	weighted results

+

78

c) Octal to Decimal

1	1	6	Octal
64	8	1	weights
64	8	6	weighted results

+

78

- From Binary to Octal or Hexadecimal

1	0 0 1	1 1 0

Group by three From right to left

Group by four from right to left

1 0 0	1 1 1 0

1 1 6

Octal

4 E

Hexadecimal

- From Octal or Hexadecimal to Binary

Octal

1 1 6

1	0 0 1	1 1 0

Binary

Hexadecimal

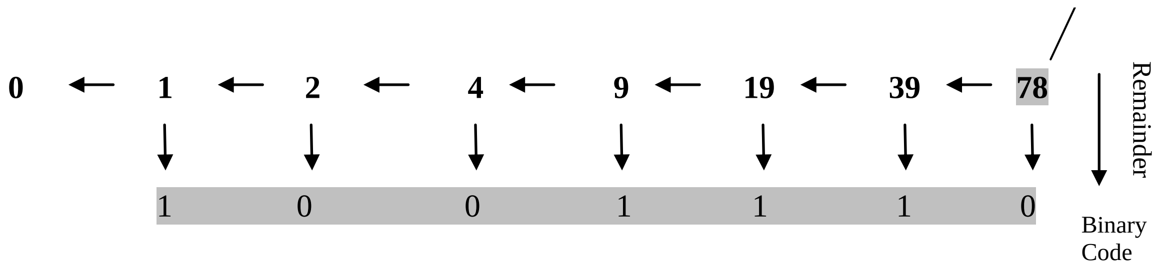
4 E

1 0 0	1 1 1 0

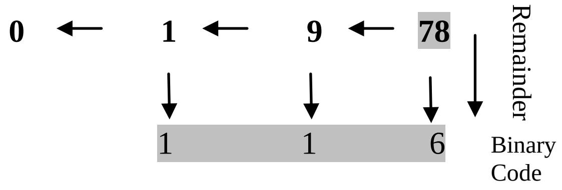
Binary

- From Decimal to Other Systems

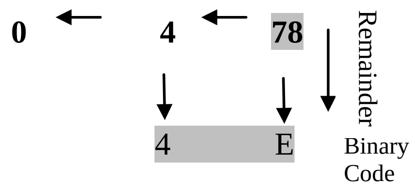
a) From Decimal to Binary (divide by 2)



b) From Decimal to Octal (divide by 8)



c) From Decimal to Hexadecimal (divide by 16)



Morse Code

Morse code is one of the oldest electrical transmission codes. The digital code system is made up of a series of dots and dashes, representing the alphabet and decimal numbers system. A dash is three times the duration of a dot.

A ∘ —
B — ∘ ∘ ∘
C — ∘ — ∘
D — ∘ ∘
.

1 ∘ — — — — —
2 ∘ ∘ — — —
.

ASCII Code

The American Standard Code for Information Interchange (ASCII) is the most widely used alphanumeric code for transmission and data processing.

ASCII is a seven-bit code that can be represented by two hexadecimal characters for simplicity. The MS hexadecimal character in this case never exceed 7

Ex

What is the ASCII code for the letter H (uppercase) in binary and hexadecimal?

Solution

Letter H is located in column 4 and row 8.

Binary: 100 1000

Hexadecimal: $ 48

Ex

What is the ASCII code for the letter k (lowercase) in binary and hexadecimal?

Solution

Letter k is located in column 6 and row B:

Binary: 110 1011

Hexadecimal: $ 6B

UNIT-5

ADVANCED MOBILE PHONE SYSTEM

- ❖ Advanced Mobile Phone Service (AMPS) is a standard system for analog signal cellular telephone service in the United States and is also used in other countries.

- ❖ It is based on the initial electromagnetic radiation spectrum allocation for cellular service by the Federal Communications Commission (FCC) in 1970. Introduced by AT&T in 1983, AMPS became one of the most widely deployed cellular system in the United States.

- ❖ AMPS allocates frequency ranges within the 800 and 900 Megahertz (MHz) spectrum to cellular telephone. Each service provider can use half of the 824-849 MHz range for receiving signals from cellular phones and half the 869-894 MHz range for transmitting to cellular phones.

- ❖ The bands are divided into 30 kHz sub-bands, called channels. The receiving channels are called reverse channels and the sending channels are called forward channels. The division of the spectrum into sub-band channels is achieved by using frequency division multiple access (FDMA).

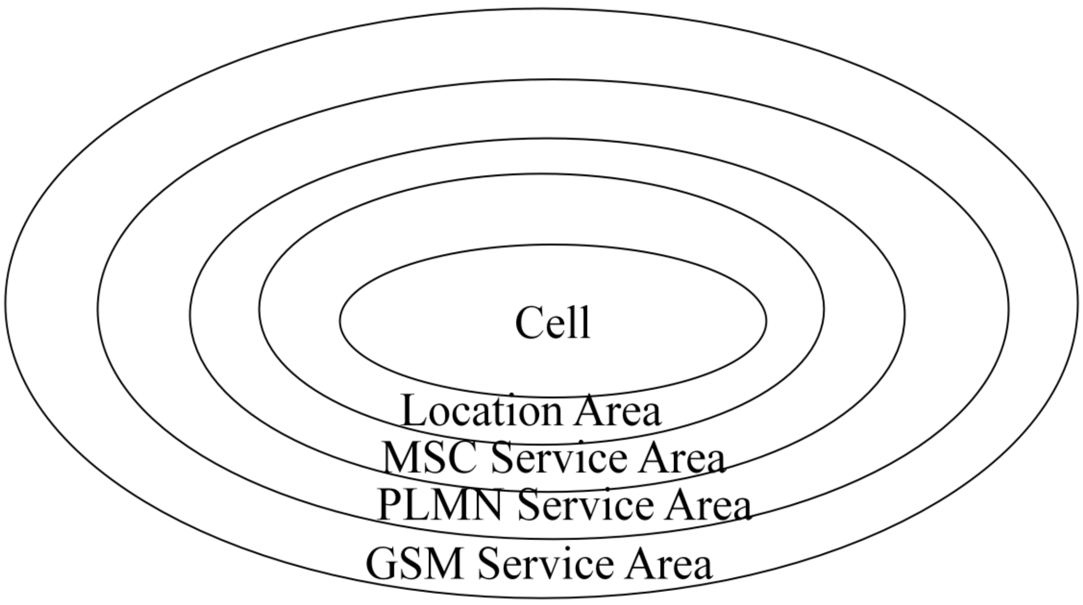
- ❖ The signals received from a transmitter cover an area called a cell. As a user moves out of the cell's area into an adjacent cell, the user begins to pick up the new cell's signals without any noticeable transition.

- The signals in the adjacent cell are sent and received on different channels than the previous cell's signals to so that the signals don't interfere with each other.

- The analog service of AMPS has been updated with digital cellular service by adding to FDMA a further subdivision of each channel using time division multiple access (TDMA). This service is known as digital AMPS (D-AMPS). Although AMPS and D-AMPS originated for the North American cellular telephone market, they are now used worldwide with over 74 million subscribers, according to Ericsson, one of the major cellular phone manufacturers.

GSM

Relation between areas in GSM



GSM MSs consist of:

- **Mobile Equipment**
- **Subscriber Identity Module**

FUNCTIONS OF MS

- Voice and data transmission & receipt
- Frequency and time synchronization
- Monitoring of power and signal quality of the surrounding cells
- Provision of location updates even during inactive state

Mobile Station

- Can receive, store, send SMS up to 160 characters.
- MS identified by unique IMEI shown on pressing *#06#.
- Power levels of 20W, 8W, 5W, 2W and .8W

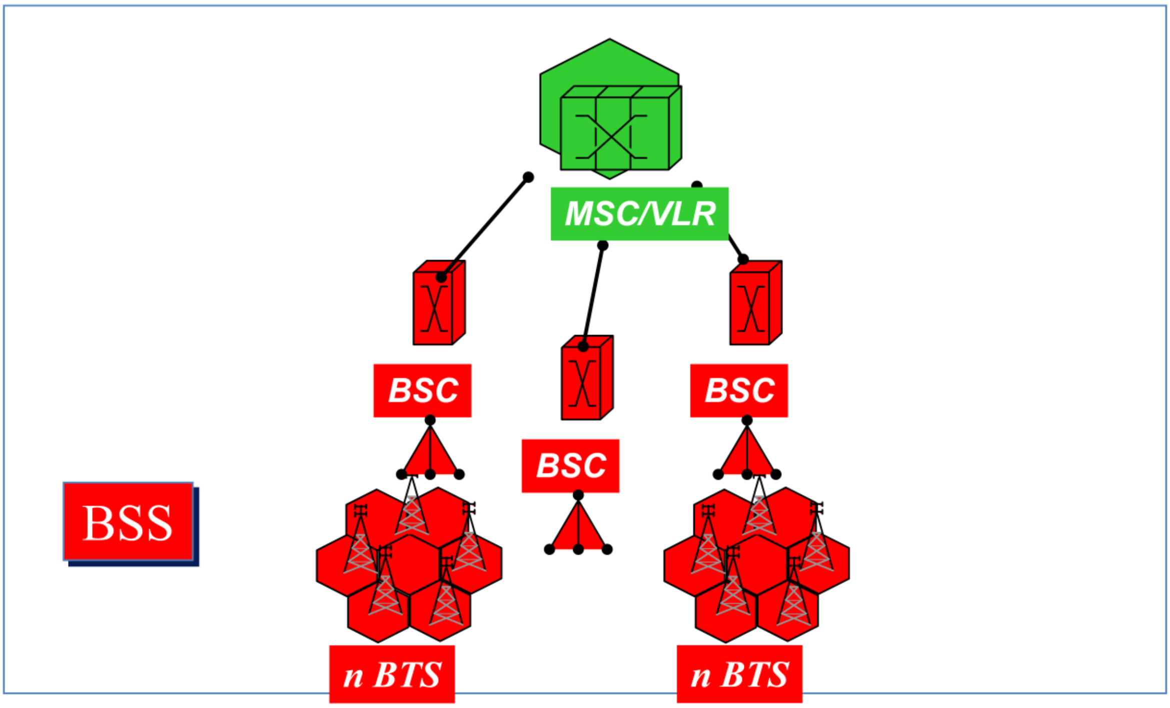
SIM

SIM has microprocessor and memory.

Fixed data stored for the subscription:

- IMSI,
- Authentication Key, Ki
- Security Algorithms:kc,A3,A8
- PIN & PUK

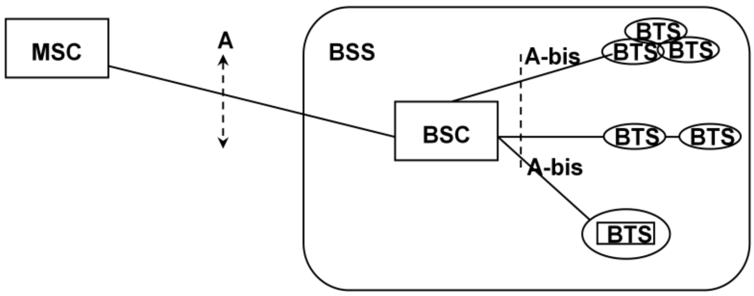
BASE STATION SYSTEM (BSS)



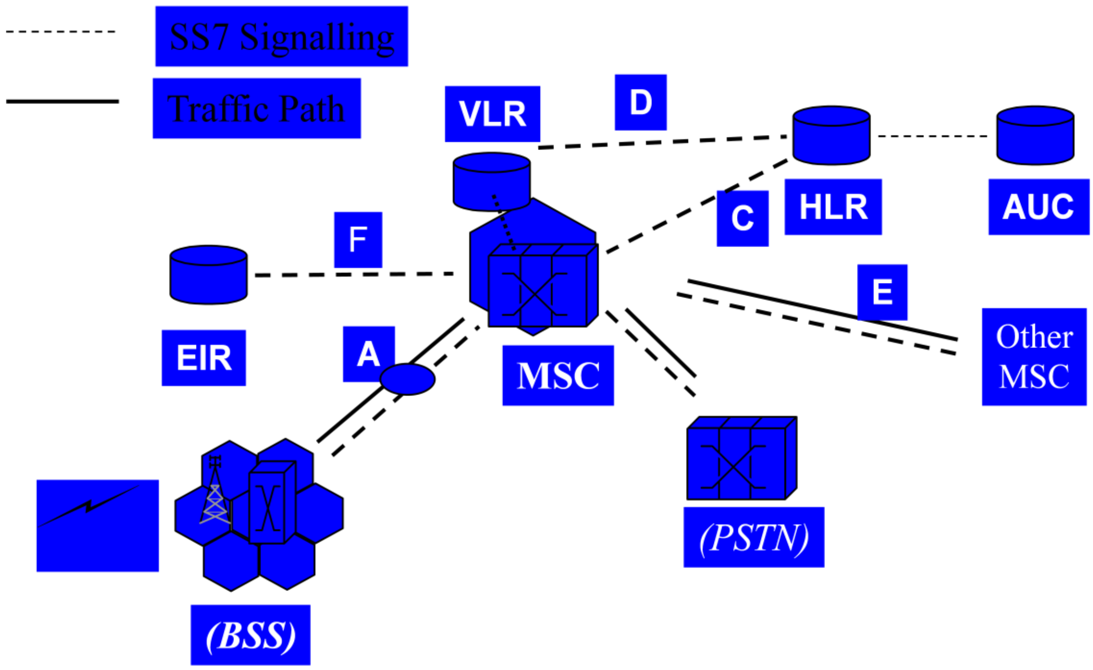
FUNCTIONS OF BTS (Base Transceiver Station)

- Radio resources
- Signal Processing
- Signaling link management
- Synchronization
- Local maintenance handling
- Functional supervision and Testing

MSC-BSS Configurations



Switching System (SS)



VISITOR LOCATION REGISTER (VLR)

- **It contains data of all mobiles roaming in its area.**

- **One VLR may be incharge of one or more LA.**

- **VLR is updated by HLR on entry of MS its area.**

- **VLR assigns TMSI which keeps on changing.**

Home Location Register(HLR)

- Reference store for subscriber's parameters, numbers, authentication & Encryption values.

- Current subscriber status and associated VLR.

- Both VLR and HLR can be implemented in the
same equipment in an MSC.

- one PLMN may contain one or several HLR.

EQUIPMENT IDENTITY REGISTER (EIR)

- This data base stores IMEI for all registered mobile equipments and is unique to every ME.

- Only one EIR per PLMN.

- *White list* : IMEI, assigned to valid ME.
- *Black list* : IMEI reported stolen
- *Gray list* : IMEI having problems like faulty software, wrong make of equipment etc.

AUthentication Center (AUC)

To authenticate the subs. attempting to use a network.

AUC is connected to HLR which provides it with authentication parameters and ciphering keys used to ensure network security.

CDMA

Code division multiple access

From Wikipedia, the free encyclopedia

This article is about a channel access method. For the mobile phone technology referred to as CDMA, see IS-95 and CDMA2000.

Multiplex techniques
Analog modulation
• AM • FM • PM • QAM • SM • SSB
Circuit mode (constant bandwidth)
• TDM • FDM / WDM • SDM • Polarization multiplexing • Spatial multiplexing • OAM multiplexing
Statistical multiplexing (variable bandwidth)
• Packet switching • Dynamic TDM • FHSS • DSSS • OFDMA

- SC-FDM
- MC-SS

Related topics

- Channel access methods
- Media access control

- v
- t
- e

Code division multiple access (**CDMA**) is a channel access method used by various radio communication technologies.

CDMA is an example of multiple access, which is where several transmitters can send information simultaneously over a single communication channel. This allows several users to share a band of frequencies (see bandwidth). To permit this to be achieved without undue interference between the users, CDMA employs spread-spectrum technology and a special coding scheme (where each transmitter is assigned a code).

CDMA is used as the access method in many mobile phone standards such as cdmaOne, CDMA2000 (the 3G evolution of cdmaOne), and WCDMA (the 3G standard used by GSM carriers), which are often referred to as simply CDMA.

ode division multiplexing (synchronous CDMA)

The digital modulation method is analogous to those used in simple radio transceivers. In the analogue case, a low frequency data signal is time multiplied with a high frequency pure sine wave carrier, and transmitted. This is effectively a frequency convolution (Weiner-Kinchin Theorem) of the two signals, resulting in a carrier with narrow sidebands. In the digital case, the sinusoidal carrier is replaced by Walsh functions. These are binary square waves that form a complete orthonormal set. The data signal is also binary and the time multiplication is achieved with a simple XOR function. This is usually a Gilbert cell mixer in the circuitry.

Synchronous CDMA exploits mathematical properties of orthogonality between vectors representing the data strings. For example, binary string 1011 is represented by the vector (1, 0, 1, 1). Vectors can be multiplied by taking their dot product, by summing the products of their respective components (for example, if u = (a, b) and v = (c, d), then their dot product u·v = ac + bd). If the dot product is zero, the two vectors are said to be orthogonal to each other. Some properties of the dot product aid understanding of how W-CDMA works. If vectors a and b are orthogonal, then $\mathbf{a}\cdot\mathbf{b}=0$and:

$$\mathbf{a}\cdot(\mathbf{a}+\mathbf{b}) = \|\mathbf{a}\|^2 \quad \text{since} \quad \mathbf{a}\cdot\mathbf{a}+\mathbf{a}\cdot\mathbf{b} = \|a\|^2+0$$

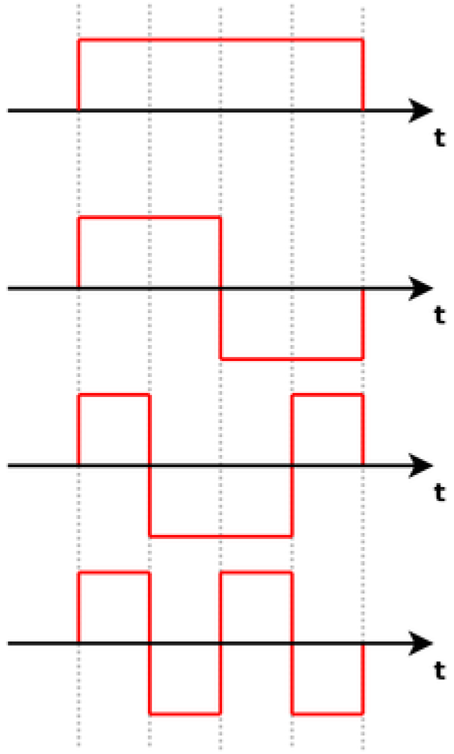
$$\mathbf{a}\cdot(-\mathbf{a}+\mathbf{b}) = -\|\mathbf{a}\|^2 \quad \text{since} \quad -\mathbf{a}\cdot\mathbf{a}+\mathbf{a}\cdot\mathbf{b} = -\|a\|^2+0$$

$$\mathbf{b}\cdot(\mathbf{a}+\mathbf{b}) = \|\mathbf{b}\|^2 \quad \text{since} \quad \mathbf{b}\cdot\mathbf{a}+\mathbf{b}\cdot\mathbf{b} = 0+\|b\|^2$$

$$\mathbf{b}\cdot(\mathbf{a}-\mathbf{b}) = -\|\mathbf{b}\|^2 \quad \text{since} \quad \mathbf{b}\cdot\mathbf{a}-\mathbf{b}\cdot\mathbf{b} = 0-\|b\|^2$$

Each user in synchronous CDMA uses a code orthogonal to the others' codes to modulate their signal. An example of four mutually orthogonal digital signals is shown in the figure. Orthogonal codes have a cross-correlation equal to zero; in other words, they do not interfere with each other. In the case of IS-95 64 bit Walsh codes are used to encode the signal to separate different users. Since each of the 64 Walsh codes are orthogonal to one another, the signals are channelized into 64 orthogonal signals. The following example demonstrates how each user's signal can be encoded and decoded.

Example



An example of four mutually orthogonal digital signals.

Start with a set of vectors that are mutually orthogonal. (Although mutual orthogonality is the only condition, these vectors are usually constructed for ease of decoding, for example columns or rows from Walsh matrices.) An example of orthogonal functions is shown in the picture on the right. These vectors will be assigned to individual users and are called the code, chip code, or chipping code. In the interest of brevity, the rest of this example uses codes, **v**, with only two bits.

Each user is associated with a different code, say **v**. A 1 bit is represented by transmitting a positive code, **v**, and a 0 bit is represented by a negative code, –**v**. For example, if $\mathbf{v} = (v_0, v_1) = (1, -1)$ and the data that the user wishes to transmit is (1, 0, 1, 1), then the transmitted symbols would be $(\mathbf{v}, -\mathbf{v}, \mathbf{v}, \mathbf{v}) = (v_0, v_1, -v_0, -v_1, v_0, v_1, v_0, v_1) = (1, -1, -1, 1, 1, -1, 1, -1)$. For the purposes of this article, we call this constructed vector the transmitted vector.

Each sender has a different, unique vector **v** chosen from that set, but the construction method of the transmitted vector is identical.

Now, due to physical properties of interference, if two signals at a point are in phase, they add to give twice the amplitude of each signal, but if they are out of phase, they subtract and give a signal that is the difference of the amplitudes. Digitally, this behaviour can be modelled by the addition of the transmission vectors, component by component.

If sender0 has code (1, –1) and data (1, 0, 1, 1), and sender1 has code (1, 1) and data (0, 0, 1, 1), and both senders transmit simultaneously, then this table describes the coding steps:

Step	Encode sender0	Encode sender1
0	code0 = (1, –1), data0 = (1, 0, 1, 1)	code1 = (1, 1), data1 = (0, 0, 1, 1)
1	encode0 = 2(1, 0, 1, 1) – (1, 1, 1, 1) = (1, –1, 1, 1)	encode1 = 2(0, 0, 1, 1) – (1, 1, 1, 1) = (–1, –1, 1, 1)
2	signal0 = encode0 ⊗ code0 = (1, –1, 1, 1) ⊗ (1, –1) = (1, –1, –1, 1, 1, –1, 1, –1)	signal1 = encode1 ⊗ code1 = (–1, –1, 1, 1) ⊗ (1, 1) = (–1, –1, –1, –1, 1, 1, 1, 1)

Because signal0 and signal1 are transmitted at the same time into the air, they add to produce the raw signal:

(1, –1, –1, 1, 1, –1, 1, –1) + (–1, –1, –1, –1, 1, 1, 1, 1) = (0, –2, –2, 0, 2, 0, 2, 0)

This raw signal is called an interference pattern. The receiver then extracts an intelligible signal for any known sender by combining the sender's code with the interference pattern, the receiver combines it with the codes of the senders. The following table explains how this works and shows that the signals do not interfere with one another:

Step	Decode sender0	Decode sender1
0	code0 = (1, –1), signal = (0, –2, –2, 0, 2, 0, 2, 0)	code1 = (1, 1), signal = (0, –2, –2, 0, 2, 0, 2, 0)
1	decode0 = pattern.vector0	decode1 = pattern.vector1
2	decode0 = ((0, –2), (–2, 0), (2, 0), (2, 0)).(1, –1)	decode1 = ((0, –2), (–2, 0), (2, 0), (2, 0)).(1, 1)
3	decode0 = ((0 + 2), (–2 + 0), (2 + 0), (2 + 0))	decode1 = ((0 – 2), (–2 + 0), (2 + 0), (2 + 0))
4	data0=(2, –2, 2, 2), meaning (1, 0, 1, 1)	data1=(–2, –2, 2, 2), meaning (0, 0, 1, 1)

Further, after decoding, all values greater than 0 are interpreted as 1 while all values less than zero are interpreted as 0. For example, after decoding, data0 is (2, –2, 2, 2), but the receiver interprets this as (1, 0, 1, 1). Values of exactly 0 means that the sender did not transmit any data, as in the following example:

Assume signal0 = (1, –1, –1, 1, 1, –1, 1, –1) is transmitted alone. The following table shows the decode at the receiver:

Step	Decode sender0	Decode sender1
0	code0 = (1, –1), signal = (1, –1, –1, 1, 1, –1, 1, –1)	code1 = (1, 1), signal = (1, –1, –1, 1, 1, –1, 1, –1)
1	decode0 = pattern.vector0	decode1 = pattern.vector1
2	decode0 = ((1, –1), (–1, 1), (1, –1), (1, –1)).(1, –1)	decode1 = ((1, –1), (–1, 1), (1, –1), (1, –1)).(1, 1)
3	decode0 = ((1 + 1), (–1 – 1),(1 +	decode1 = ((1 – 1), (–1 + 1),(1

	1), (1 + 1))	– 1), (1 – 1))
4	data0 = (2, –2, 2, 2), meaning (1, 0, 1, 1)	data1 = (0, 0, 0, 0), meaning no data

When the receiver attempts to decode the signal using sender1's code, the data is all zeros, therefore the cross correlation is equal to zero and it is clear that sender1 did not transmit any data.

Asynchronous CDMA

See also: Direct-sequence spread spectrum and near-far problem

When mobile-to-base links cannot be precisely coordinated, particularly due to the mobility of the handsets, a different approach is required. Since it is not mathematically possible to create signature sequences that are both orthogonal for arbitrarily random starting points and which make full use of the code space, unique "pseudo-random" or "pseudo-noise" (PN) sequences are used in asynchronous CDMA systems. A PN code is a binary sequence that appears random but can be reproduced in a deterministic manner by intended receivers. These PN codes are used to encode and decode a user's signal in Asynchronous CDMA in the same manner as the orthogonal codes in synchronous CDMA (shown in the example above). These PN sequences are statistically uncorrelated, and the sum of a large number of PN sequences results in multiple access interference (MAI) that is approximated by a Gaussian noise process (following the central limit theorem in statistics). Gold codes are an example of a PN suitable for this purpose, as there is low correlation between the codes. If all of the users are received with the same power level, then the variance (e.g., the noise power) of the MAI increases in direct proportion to the number of users. In other words, unlike synchronous CDMA, the signals of other users will appear as noise to the signal of interest and interfere slightly with the desired signal in proportion to number of users.

All forms of CDMA use spread spectrum process gain to allow receivers to partially discriminate against unwanted signals. Signals encoded with the specified PN sequence (code) are received, while signals with different codes (or the same code but a different timing offset) appear as wideband noise reduced by the process gain.

Since each user generates MAI, controlling the signal strength is an important issue with CDMA transmitters. A CDM (synchronous CDMA), TDMA, or FDMA receiver can in theory completely reject arbitrarily strong signals using different codes, time slots or frequency channels due to the orthogonality of these systems. This is not true for Asynchronous CDMA; rejection of unwanted signals is only partial. If any or all of the unwanted signals are much stronger than the desired signal, they will overwhelm it. This leads to a general requirement in any asynchronous CDMA system to approximately match the various signal power levels as seen at the receiver. In CDMA cellular, the base station uses a fast closed-loop power control scheme to tightly control each mobile's transmit power.

Advantages of asynchronous CDMA over other techniques

Efficient practical utilization of the fixed frequency spectrum

In theory CDMA, TDMA and FDMA have exactly the same spectral efficiency but practically, each has its own challenges – power control in the case of CDMA, timing in the case of TDMA, and frequency generation/filtering in the case of FDMA.

TDMA systems must carefully synchronize the transmission times of all the users to ensure that they are received in the correct time slot and do not cause interference. Since this cannot be perfectly controlled in a mobile environment, each time slot must have a guard-time, which reduces the probability that users will interfere, but decreases the spectral efficiency. Similarly, FDMA systems must use a guard-band between adjacent channels, due to the unpredictable doppler shift of the signal spectrum because of user mobility. The guard-bands will reduce the probability that adjacent channels will interfere, but decrease the utilization of the spectrum.

Flexible allocation of resources

Asynchronous CDMA offers a key advantage in the flexible allocation of resources i.e. allocation of a PN codes to active users. In the case of CDM (synchronous CDMA), TDMA, and FDMA the number of simultaneous orthogonal codes, time slots and frequency slots respectively are fixed hence the capacity in terms of number of simultaneous users is limited. There are a fixed number of orthogonal codes, time slots or frequency bands that can be allocated for CDM, TDMA, and FDMA systems, which remain underutilized due to the bursty nature of telephony and packetized data transmissions. There is no strict limit to the number of users that can be supported in an asynchronous CDMA system, only a practical limit governed by the desired bit error probability, since the SIR (Signal to Interference Ratio) varies inversely with the number of users. In a bursty traffic environment like mobile telephony, the advantage afforded by asynchronous CDMA is that the performance (bit error rate) is allowed to fluctuate randomly, with an average value determined by the number of users times the percentage of utilization. Suppose there are 2N users that only talk half of the time, then 2N users can be accommodated with the same average bit error probability as N users that talk all of the time. The key difference here is that the bit error probability for N users talking all of the time is constant, whereas it is a random quantity (with the same mean) for 2N users talking half of the time.

In other words, asynchronous CDMA is ideally suited to a mobile network where large numbers of transmitters each generate a relatively small amount of traffic at irregular intervals. CDM (synchronous CDMA), TDMA, and FDMA systems cannot recover the underutilized resources inherent to bursty traffic due to the fixed number of orthogonal codes, time slots or frequency channels that can be assigned to individual transmitters. For instance, if there are N time slots in a TDMA system and 2N users that talk half of the time, then half of the time there will be more than N users needing to use more than N time slots. Furthermore, it would require significant overhead to continually allocate and deallocate the orthogonal code, time slot or frequency channel resources. By comparison, asynchronous CDMA transmitters simply send when they have something to say, and go off the air when they don't, keeping the same PN signature sequence as long as they are connected to the system.

Spread-spectrum characteristics of CDMA

Most modulation schemes try to minimize the bandwidth of this signal since bandwidth is a limited resource. However, spread spectrum techniques use a transmission bandwidth that is several orders of magnitude greater than the minimum required signal bandwidth. One of the initial reasons for doing this was military applications including guidance and communication systems. These systems were designed using spread spectrum because of its security and resistance to jamming. Asynchronous CDMA has some level of privacy built in because the signal is spread using a pseudo-random code; this code makes the spread spectrum signals appear random or have noise-like properties. A receiver cannot demodulate this transmission without knowledge of the pseudo-random sequence used to encode the data. CDMA is also resistant to jamming. A jamming signal only has a finite amount of power available to jam the signal. The jammer can either spread its energy over the entire bandwidth of the signal or jam only part of the entire signal.[9]

CDMA can also effectively reject narrow band interference. Since narrow band interference affects only a small portion of the spread spectrum signal, it can easily be removed through notch filtering without much loss of information. Convolution encoding and interleaving can be used to assist in recovering this lost data. CDMA signals are also resistant to multipath fading. Since the spread spectrum signal occupies a large bandwidth only a small portion of this will undergo fading due to multipath at any given time. Like the narrow band interference this will result in only a small loss of data and can be overcome.

Another reason CDMA is resistant to multipath interference is because the delayed versions of the transmitted pseudo-random codes will have poor correlation with the original pseudo-random code, and will thus appear as another user, which is ignored at the receiver. In other words, as long as the multipath channel induces at least one chip of delay, the multipath signals will arrive at the receiver such that they are shifted in time by at least one chip from the intended signal. The correlation properties of the pseudo-random codes are such that this slight delay causes the multipath to appear uncorrelated with the intended signal, and it is thus ignored.

Some CDMA devices use a rake receiver, which exploits multipath delay components to improve the performance of the system. A rake receiver combines the information from several correlators, each one tuned to a different path delay, producing a stronger version of the signal than a simple receiver with a single correlation tuned to the path delay of the strongest signal.[10]

Frequency reuse is the ability to reuse the same radio channel frequency at other cell sites within a cellular system. In the FDMA and TDMA systems frequency planning is an important consideration. The frequencies used in different cells must be planned carefully to ensure signals from different cells do not interfere with each other. In a CDMA system, the same frequency can be used in every cell, because channelization is done using the pseudo-random codes. Reusing the same frequency in every cell eliminates the need for frequency planning in a CDMA system; however, planning of the different pseudo-random sequences must be done to ensure that the received signal from one cell does not correlate with the signal from a nearby cell.[11]

Since adjacent cells use the same frequencies, CDMA systems have the ability to perform soft hand offs. Soft hand offs allow the mobile telephone to communicate simultaneously with two or more cells. The best signal quality is selected until the hand off is complete. This is different from hard hand offs utilized in other cellular systems. In a hard hand off situation, as the mobile telephone approaches a hand off, signal strength may vary abruptly. In contrast, CDMA systems use the soft hand off, which is undetectable and provides a more reliable and higher quality signal.[11]

Collaborative CDMA

In a recent study, a novel collaborative multi-user transmission and detection scheme called Collaborative CDMA[12] has been investigated for the uplink that exploits the differences between users' fading channel signatures to increase the user capacity well beyond the spreading length in multiple access interference (MAI) limited environment. The authors show that it is possible to achieve this increase at a low complexity and high bit error rate performance in flat fading channels, which is a major research challenge for overloaded CDMA systems. In this approach, instead of using one sequence per user as in conventional CDMA, the authors group a small number of users to share the same spreading sequence and enable group spreading and despreading operations. The new collaborative multi-user receiver consists of two stages: group multi-user detection (MUD) stage to suppress the MAI between the groups and a low complexity maximum-likelihood detection stage to recover jointly the co-spread users' data using minimum Euclidean distance measure and users' channel gain coefficients. In CDM signal security is high.

HAND OFF

In cellular telecommunications, the term **handover** or **handoff** refers to the process of transferring an ongoing call or data session from one channel connected to the core network to another channel. In satellite communications it is the process of transferring satellite control responsibility from one earth station to another without loss or interruption of service.

Purpose

In telecommunications there may be different reasons why a handover might be conducted:

- when the phone is moving away from the area covered by one cell and entering the area covered by another cell the call is transferred to the second cell in order to avoid call termination when the phone gets outside the range of the first cell;

- when the capacity for connecting new calls of a given cell is used up and an existing or new call from a phone, which is located in an area overlapped by another cell, is transferred to that cell in order to free-up some capacity in the first cell for other users, who can only be connected to that cell;
- in non-CDMA networks when the channel used by the phone becomes interfered by another phone using the same channel in a different cell, the call is transferred to a different channel in the same cell or to a different channel in another cell in order to avoid the interference;
- again in non-CDMA networks when the user behaviour changes, e.g. when a fast-travelling user, connected to a large, umbrella-type of cell, stops then the call may be transferred to a smaller macro cell or even to a micro cell in order to free capacity on the umbrella cell for other fast-traveling users and to reduce the potential interference to other cells or users (this works in reverse too, when a user is detected to be moving faster than a certain threshold, the call can be transferred to a larger umbrella-type of cell in order to minimize the frequency of the handovers due to this movement);
- in CDMA networks a handover (see further down) may be induced in order to reduce the interference to a smaller neighboring cell due to the "near-far" effect even when the phone still has an excellent connection to its current cell;
- etc.

The most basic form of handover is when a phone call in progress is redirected from its current cell (called source) to a new cell (called target). In terrestrial networks the source and the target cells may be served from two different cell sites or from one and the same cell site (in the latter case the two cells are usually referred to as two sectors on that cell site). Such a handover, in which the source and the target are different cells (even if they are on the same cell site) is called inter-cell handover. The purpose of inter-cell handover is to maintain the call as the subscriber is moving out of the area covered by the source cell and entering the area of the target cell.

A special case is possible, in which the source and the target are one and the same cell and only the used channel is changed during the handover. Such a handover, in which the cell is not changed, is called intra-cell handover. The purpose of intra-cell handover is to change one channel, which may be interfered or fading with a new clearer or less fading channel.

Types of handover

In addition to the above classification of inter-cell and intra-cell classification of handovers, they also can be divided into hard and soft handovers:

- A hard handover is one in which the channel in the source cell is released and only then the channel in the target cell is engaged. Thus the connection to the source is broken before or 'as' the connection to the target is made—for this reason such handovers are also known as break-before-make. Hard handovers are intended to be instantaneous in order to minimize the disruption to the call. A hard handover is perceived by network engineers as an event during the call. It requires the least processing by the network providing service. When the mobile is between base stations, then the mobile can switch with any of the base stations, so the base stations bounce the link with the mobile back and forth. This is called ping-ponging.
- A soft handover is one in which the channel in the source cell is retained and used for a while in parallel with the channel in the target cell. In this case the connection to the target is established before the connection to the source is broken, hence this handover is called make-before-break. The interval, during which the two connections are used in parallel, may be brief or substantial. For this reason the soft handover is perceived by network engineers as a state of the call, rather than a brief event. Soft handovers may involve using connections to more than two cells: connections to three, four or more cells can be maintained by one phone at the same time. When a call is in a state of soft handover, the signal of the best of all used channels can be used for the call at a given moment or all the signals can be combined

to produce a clearer copy of the signal. The latter is more advantageous, and when such combining is performed both in the downlink (forward link) and the uplink (reverse link) the handover is termed as softer. Softer handovers are possible when the cells involved in the handovers have a single cell site.

Comparison of handovers

An advantage of the hard handover is that at any moment in time one call uses only one channel. The hard handover event is indeed very short and usually is not perceptible by the user. In the old analog systems it could be heard as a click or a very short beep; in digital systems it is unnoticeable. Another advantage of the hard handoff is that the phone's hardware does not need to be capable of receiving two or more channels in parallel, which makes it cheaper and simpler. A disadvantage is that if a handover fails the call may be temporarily disrupted or even terminated abnormally. Technologies which use hard handovers, usually have procedures which can re-establish the connection to the source cell if the connection to the target cell cannot be made. However re-establishing this connection may not always be possible (in which case the call will be terminated) and even when possible the procedure may cause a temporary interruption to the call.

One advantage of the soft handovers is that the connection to the source cell is broken only when a reliable connection to the target cell has been established and therefore the chances that the call will be terminated abnormally due to failed handovers are lower. However, by far a bigger advantage comes from the mere fact that simultaneously channels in multiple cells are maintained and the call could only fail if all of the channels are interfered or fade at the same time. Fading and interference in different channels are unrelated and therefore the probability of them taking place at the same moment in all channels is very low. Thus the reliability of the connection becomes higher when the call is in a soft handover. Because in a cellular network the majority of the handovers occur in places of poor coverage, where calls would frequently become unreliable when their channel is interfered or fading, soft handovers bring a significant improvement to the reliability of the calls in these places by making the interference or the fading in a single channel not critical. This advantage comes at the cost of more complex hardware in the phone, which must be capable of processing several channels in parallel. Another price to pay for soft handovers is use of several channels in the network to support just a single call. This reduces the number of remaining free channels and thus reduces the capacity of the network. By adjusting the duration of soft handovers and the size of the areas in which they occur, the network engineers can balance the benefit of extra call reliability against the price of reduced capacity.

Possibility of handover

While theoretically speaking soft handovers are possible in any technology, analog or digital, the cost of implementing them for analog technologies is prohibitively high and none of the technologies that were commercially successful in the past (e.g. AMPS, TACS, NMT, etc.) had this feature. Of the digital technologies, those based on FDMA also face a higher cost for the phones (due to the need to have multiple parallel radio-frequency modules) and those based on TDMA or a combination of TDMA/FDMA, in principle, allow not so expensive implementation of soft handovers. However, none of the 2G (second-generation) technologies have this feature (e.g. GSM, D-AMPS/IS-136, etc.). On the other hand, all CDMA based technologies, 2G and 3G (third-generation), have soft handovers. On one hand, this is facilitated by the possibility to design not so expensive phone hardware supporting soft handovers for CDMA and on the other hand, this is necessitated by the fact that without soft handovers CDMA networks may suffer from substantial interference arising due to the so-called near-far effect.

In all current commercial technologies based on FDMA or on a combination of TDMA/FDMA (e.g. GSM, AMPS, IS-136/DAMPS, etc.) changing the channel during a hard handover is realised by changing the pair of used transmit/receive frequencies.

Implementations

For the practical realisation of handoffs in a cellular network each cell is assigned a list of potential target cells, which can be used for handing-off calls from this source cell to them. These potential target cells are called neighbours and the list is called neighbour list. Creating such a list for a given cell is not trivial and specialised computer tools are used. They implement different algorithms and may use for input data from field measurements or computer predictions of radio wave propagation in the areas covered by the cells.

During a call one or more parameters of the signal in the channel in the source cell are monitored and assessed in order to decide when a handover may be necessary. The downlink (forward link) and/or uplink (reverse link) directions may be monitored. The handover may be requested by the phone or by the base station (BTS) of its source cell and, in some systems, by a BTS of a neighbouring cell. The phone and the BTSs of the neighbouring cells monitor each other others' signals and the best target candidates are selected among the neighbouring cells. In some systems, mainly based on CDMA, a target candidate may be selected among the cells which are not in the neighbour list. This is done in an effort to reduce the probability of interference due to the aforementioned near-far effect.

In analog systems the parameters used as criteria for requesting a hard handover are usually the received signal power and the received signal-to-noise ratio (the latter may be estimated in an analog system by inserting additional tones, with frequencies just outside the captured voice-frequency band at the transmitter and assessing the form of these tones at the receiver). In non-CDMA 2G digital systems the criteria for requesting hard handover may be based on estimates of the received signal power, bit error rate (BER) and block error/erasure rate (BLER), received quality of speech (RxQual), distance between the phone and the BTS (estimated from the radio signal propagation delay) and others. In CDMA systems, 2G and 3G, the most common criterion for requesting a handover is Ec/Io ratio measured in the pilot channel (CPICH) and/or RSCP.

In CDMA systems, when the phone in soft or softer handoff is connected to several cells simultaneously, it processes the received in parallel signals using a rake receiver. Each signal is processed by a module called rake finger. A usual design of a rake receiver in mobile phones includes three or more rake fingers used in soft handoff state for processing signals from as many cells and one additional finger used to search for signals from other cells. The set of cells, whose signals are used during a soft handoff, is referred to as the active set. If the search finger finds a sufficiently-strong signal (in terms of high Ec/Io or RSCP) from a new cell this cell is added to the active set. The cells in the neighbour list (called in CDMA neighbouring set) are checked more frequently than the rest and thus a handoff with a neighbouring cell is more likely, however a handoff with others cells outside the neighbor list is also allowed (unlike in GSM, IS-136/DAMPS, AMPS, NMT, etc.).

Satellite

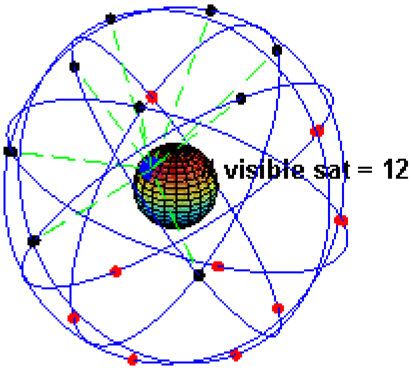
From Wikipedia, the free encyclopedia

This article is about artificial satellites. For natural satellites, also known as moons, see Natural satellite. For other uses, see Satellite (disambiguation).

This article is **outdated**. Please update this article to reflect recent events or newly available information. (December 2013)

Play media

NASA's Earth-observing fleet as of June 2012.



An animation depicting the orbits of GPS satellites in medium Earth orbit.

A full-size model of the Earth observation satellite ERS 2

In the context of spaceflight, a **satellite** is an artificial object which has been intentionally placed into orbit. Such objects are sometimes called **artificial satellites** to distinguish them from natural satellites such as the Moon.

The world's first artificial satellite, the Sputnik 1, was launched by the Soviet Union in 1957. Since then, thousands of satellites have been launched into orbit around the Earth. Some satellites, notably space stations, have been launched in parts and assembled in orbit. Artificial satellites originate from more than 50 countries and have used the satellite launching capabilities of ten nations. A few hundred satellites are currently operational, whereas thousands of unused satellites and satellite fragments orbit the Earth as space debris. A few space probes have been placed into orbit around other bodies and become artificial satellites to the Moon, Mercury, Venus, Mars, Jupiter, Saturn, Vesta, Eros, and the Sun.

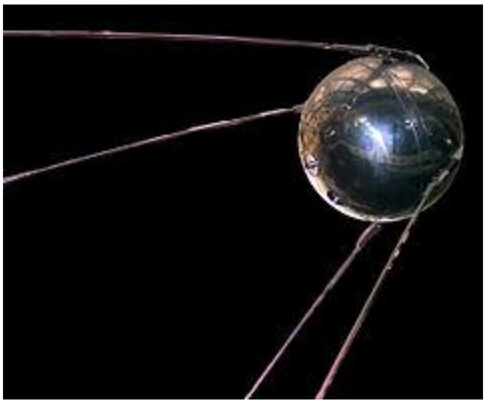
Satellites are used for a large number of purposes. Common types include military and civilian Earth observation satellites, communications satellites, navigation satellites, weather satellites, and research satellites. Space stations and human spacecraft in orbit are also satellites. Satellite orbits vary greatly, depending on the purpose of the satellite, and are classified in a number of ways. Well-known (overlapping) classes include low Earth orbit, polar orbit, and geostationary orbit.

About 6,600 satellites have been launched. The latest estimates are that 3,600 remain in orbit.[1] Of those, about 1,000 are operational;[2][3] the rest have lived out their useful lives and are part of the space debris. Approximately 500 operational satellites are in low-Earth orbit, 50 are in medium-Earth orbit (at 20,000 km), the rest are in geostationary orbit (at 36,000 km).[4]

Satellites are propelled by rockets to their orbits. Usually the launch vehicle itself is a rocket lifting off from a launch pad on land. In a minority of cases satellites are launched at sea (from a submarine or a mobile maritime platform) or aboard a plane (see air launch to orbit).

Satellites are usually semi-independent computer-controlled systems. Satellite subsystems attend many tasks, such as power generation, thermal control, telemetry, attitude control and orbit control.

History of artificial satellites



Sputnik 1: The first artificial satellite to orbit Earth.

The first artificial satellite was Sputnik 1, launched by the Soviet Union on October 4, 1957, and initiating the Soviet Sputnik program, with Sergei Korolev as chief designer (there is a crater on the lunar far side which bears his name). This in turn triggered the Space Race between the Soviet Union and the United States.

Sputnik 1 helped to identify the density of high atmospheric layers through measurement of its orbital change and provided data on radio-signal distribution in the ionosphere. The unanticipated announcement of Sputnik 1's success precipitated the Sputnik crisis in the United States and ignited the so-called Space Race within the Cold War.

Sputnik 2 was launched on November 3, 1957 and carried the first living passenger into orbit, a dog named Laika.[9]

In May, 1946, Project RAND had released the Preliminary Design of an Experimental World-Circling Spaceship, which stated, "A satellite vehicle with appropriate instrumentation can be expected to be one of the most potent scientific tools of the Twentieth Century."[10] The United States had been considering launching orbital satellites since

1945 under the Bureau of Aeronautics of the United States Navy. The United States Air Force's Project RAND eventually released the above report, but did not believe that the satellite was a potential military weapon; rather, they considered it to be a tool for science, politics, and propaganda. In 1954, the Secretary of Defense stated, "I know of no American satellite program."[11]

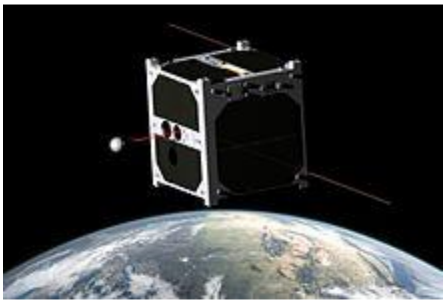
On July 29, 1955, the White House announced that the U.S. intended to launch satellites by the spring of 1958. This became known as Project Vanguard. On July 31, the Soviets announced that they intended to launch a satellite by the fall of 1957.

Following pressure by the American Rocket Society, the National Science Foundation, and the International Geophysical Year, military interest picked up and in early 1955 the Army and Navy were working on Project Orbiter, two competing programs: the army's which involved using a Jupiter C rocket, and the civilian/Navy Vanguard Rocket, to launch a satellite. At first, they failed: initial preference was given to the Vanguard program, whose first attempt at orbiting a satellite resulted in the explosion of the launch vehicle on national television. But finally, three months after Sputnik 2, the project succeeded; Explorer 1 became the United States' first artificial satellite on January 31, 1958.[12]

In June 1961, three-and-a-half years after the launch of Sputnik 1, the Air Force used resources of the United States Space Surveillance Network to catalog 115 Earth-orbiting satellites.[13]

Early satellites were constructed as "one-off" designs. With growth in geosynchronous (GEO) satellite communication, multiple satellites began to be built on single model platforms called satellite buses. The first standardized satellite bus design was the HS-333 GEO commsat, launched in 1972.

The largest artificial satellite currently orbiting the Earth is the International Space Station.



1U CubeSat ESTCube-1, developed mainly by the students from the University of Tartu, carries out a tether deployment experiment on the low Earth orbit.

Space Surveillance Network

Main article: United States Space Surveillance Network

The United States Space Surveillance Network (SSN), a division of The United States Strategic Command, has been tracking objects in Earth's orbit since 1957 when the Soviets opened the space age with the launch of Sputnik I. Since then, the SSN has tracked more than 26,000 objects. The SSN currently tracks more than 8,000 man-made orbiting objects. The rest have re-entered Earth's atmosphere and disintegrated, or survived re-entry and impacted the Earth. The SSN tracks objects that are 10 centimeters in diameter or larger; those now orbiting Earth range from satellites weighing several tons to pieces of spent rocket bodies weighing only 10 pounds. About seven percent are operational satellites (i.e. ~560 satellites), the rest are space debris.[14] The United States Strategic Command is primarily interested in the active satellites, but also tracks space debris which upon reentry might otherwise be mistaken for incoming missiles.

A search of the NSSDC Master Catalog at the end of October 2010 listed 6,578 satellites launched into orbit since 1957, the latest being Chang'e 2, on 1 October 2010.[15]

Non-military satellite services

There are three basic categories of non-military satellite services:[16]

Fixed satellite services

Fixed satellite services handle hundreds of billions of voice, data, and video transmission tasks across all countries and continents between certain points on the Earth's surface.

Mobile satellite systems

Mobile satellite systems help connect remote regions, vehicles, ships, people and aircraft to other parts of the world and/or other mobile or stationary communications units, in addition to serving as navigation systems.

Scientific research satellites (commercial and noncommercial)

Scientific research satellites provide meteorological information, land survey data (e.g. remote sensing), Amateur (HAM) Radio, and other different scientific research applications such as earth science, marine science, and atmospheric research.

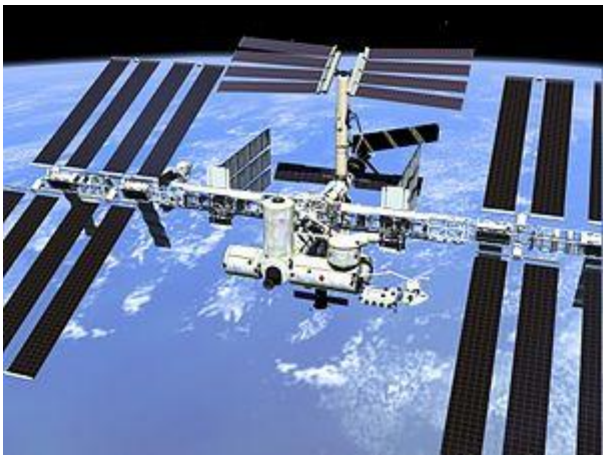
Types

MILSTAR: A communication satellite

- **"Killer Satellites"** are satellites that are designed to destroy enemy warheads, satellites, and other space assets.
- **Astronomical satellites** are satellites used for observation of distant planets, galaxies, and other outer space objects.

- **Biosatellites** are satellites designed to carry living organisms, generally for scientific experimentation.
- **Communications satellites** are satellites stationed in space for the purpose of telecommunications. Modern communications satellites typically use geosynchronous orbits, Molniya orbits or Low Earth orbits.
- **Miniaturized satellites** are satellites of unusually low masses and small sizes.[17] New classifications are used to categorize these satellites: minisatellite (500–100 kg), microsatellite (below 100 kg), nanosatellite (below 10 kg).[citation needed]
- **Navigational satellites** are satellites which use radio time signals transmitted to enable mobile receivers on the ground to determine their exact location. The relatively clear line of sight between the satellites and receivers on the ground, combined with ever-improving electronics, allows satellite navigation systems to measure location to accuracies on the order of a few meters in real time.
- **Reconnaissance satellites** are Earth observation satellite or communications satellite deployed for military or intelligence applications. Very little is known about the full power of these satellites, as governments who operate them usually keep information pertaining to their reconnaissance satellites classified.
- **Earth observation satellites** are satellites intended for non-military uses such as environmental monitoring, meteorology, map making etc. (See especially Earth Observing System.)
- **Tether satellites** are satellites which are connected to another satellite by a thin cable called a tether.
- **Weather satellites** are primarily used to monitor Earth's weather and climate.[18]
- **Recovery satellites** are satellites that provide a recovery of reconnaissance, biological, space-production and other payloads from orbit to Earth.
- **Manned spacecraft (spaceships)** are large satellites able to put humans into (and beyond) an orbit, and return them to Earth. Spacecraft including spaceplanes of reusable systems have

major propulsion or landing facilities. They can be used as transport to and from the orbital stations.



International Space Station as seen from Space

- **Space stations** are man-made orbital structures that are designed for human beings to live on in outer space. A space station is distinguished from other manned spacecraft by its lack of major propulsion or landing facilities. Space stations are designed for medium-term living in orbit, for periods of weeks, months, or even years.
- A **Skyhook** is a proposed type of tethered satellite/ion powered space station that serves as a terminal for suborbital launch vehicles flying between the Earth and the lower end of the Skyhook, as well as a terminal for spacecraft going to, or arriving from, higher orbit, the Moon, or Mars, at the upper end of the Skyhook.[19][20]

BLUETOOTH

This article is about a wireless technology standard. For the medieval King of Denmark, see Harald Bluetooth.

Bluetooth

Bluetooth

Developed by	Bluetooth Special Interest Group
Industry	Mobile personal area networks
Compatible hardware	Mobile phones, Personal computers, Laptop computers
Physical range	Up to 60 metres[1]

Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength UHF radio waves in the ISM band from 2.4 to 2.485 GHz[2]) from fixed and mobile devices, and building personal area networks (PANs). Invented by telecom vendor Ericsson in 1994,[3] it was originally conceived as a wireless alternative to RS-232 data cables. It can connect several devices, overcoming problems of synchronization.

Bluetooth is managed by the Bluetooth Special Interest Group (SIG), which has more than 20,000 member companies in the areas of telecommunication, computing, networking, and consumer electronics.[4] Bluetooth was standardized as **IEEE 802.15.1**, but the standard is no longer maintained. The SIG oversees the development of the specification, manages the qualification program, and protects the trademarks.[5] To be marketed as a Bluetooth device, it must be qualified to standards defined by the SIG.[6] A network of patents is required to implement the technology, which is licensed only for that qualifying device.

Implementation

Bluetooth operates in the range of 2400–2483.5 MHz (including guard bands). This is in the globally unlicensed (but not unregulated) Industrial, Scientific and Medical (ISM) 2.4 GHz short-range radio frequency band. Bluetooth uses a radio technology called frequency-hopping spread spectrum. The transmitted data are divided into packets and each packet is transmitted on one of the 79 designated Bluetooth channels. Each channel has a bandwidth of 1 MHz. Bluetooth 4.0 uses 2 MHz spacing which allows for 40 channels. The first channel starts at 2402 MHz and continues up to 2480 MHz in 1 MHz steps. It usually performs 1600 hops per second, with Adaptive Frequency-Hopping (AFH) enabled.[12]

Originally, Gaussian frequency-shift keying (GFSK) modulation was the only modulation scheme available; subsequently, since the introduction of Bluetooth 2.0+EDR, π/4-DQPSK and 8DPSK modulation may also be used between compatible devices. Devices functioning with GFSK are said to be operating in basic rate (BR) mode where an instantaneous data rate of 1 Mbit/s is possible. The term Enhanced Data Rate (EDR) is used to describe π/4-DPSK and 8DPSK schemes, each giving 2 and 3 Mbit/s respectively. The combination of these (BR and EDR) modes in Bluetooth radio technology is classified as a "BR/EDR radio".

Bluetooth is a packet-based protocol with a master-slave structure. One master may communicate with up to seven slaves in a piconet; all devices share the master's clock. Packet exchange is based on the basic clock, defined by the master, which ticks at 312.5 µs intervals. Two clock ticks make up a slot of 625 µs; two slots make up a slot pair of 1250 µs. In the simple case of single-slot packets the master transmits in even slots and receives in odd slots; the slave, conversely, receives in even slots and transmits in odd slots. Packets may be 1, 3 or 5 slots long, but in all cases the master transmit will begin in even slots and the slave transmit in odd slots.

Communication and connection

A master Bluetooth device can communicate with a maximum of seven devices in a piconet (an ad-hoc computer network using Bluetooth technology), though not all devices reach this maximum. The devices can switch roles, by agreement, and the slave can become the master (for example, a headset initiating a connection to a phone will necessarily begin as master, as initiator of the connection; but may subsequently prefer to be slave).

The Bluetooth Core Specification provides for the connection of two or more piconets to form a scatternet, in which certain devices simultaneously play the master role in one piconet and the slave role in another.

At any given time, data can be transferred between the master and one other device (except for the little-used broadcast mode.[citation needed]) The master chooses which slave device to address; typically, it switches rapidly from one device to another in a round-robin fashion. Since it is the master that chooses which slave to address, whereas a slave is (in theory) supposed to listen in each receive slot, being a master is a lighter burden than being a slave. Being a master of seven slaves is possible; being a slave of more than one master is difficult.[citation needed] The specification is vague as to required behavior in scatternets.

Many USB Bluetooth adapters or "dongles" are available, some of which also include an IrDA adapter.[citation needed]

Uses

Class	Max. permitted power (mW)	(dBm)	Typ. range[13] (m)
1	100	20	~100
2	2.5	4	~10

3	1	0	~1

Bluetooth is a standard wire-replacement communications protocol primarily designed for low-power consumption, with a short range based on low-cost transceiver microchips in each device.[14] Because the devices use a radio (broadcast) communications system, they do not have to be in visual line of sight of each other, however a quasi optical wireless path must be viable.[4] Range is power-class-dependent, but effective ranges vary in practice; see the table on the right.

Version	Data rate	Max. application throughput
1.2	1 Mbit/s	>80 kbit/s
2.0 + EDR	3 Mbit/s	>80 kbit/s
3.0 + HS	24 Mbit/s	See Version 3.0 + HS
4.0	24 Mbit/s	See Version 4.0 LE

The effective range varies due to propagation conditions, material coverage, production sample variations, antenna configurations and battery conditions. Most Bluetooth applications are in indoor conditions, where attenuation of walls and signal fading due to signal reflections will cause the range to be far lower than the specified line-of-sight ranges of the Bluetooth products. Most Bluetooth applications are battery powered Class 2 devices, with little difference in range whether the other end of the link is a Class 1 or Class 2 device as the lower powered device tends to set the range limit. In some cases the effective range of the data link can be extended when a Class 2 devices is connecting to a Class 1 transceiver with both higher sensitivity and transmission power than a typical Class 2 device.[15] Mostly however the Class 1 devices have a similar sensitivity to Class 2 devices. Connecting two Class 1 devices with both high sensitivity and high power can allow ranges far in excess of the typical 100m, depending on the throughput required by the application. Some such devices allow open field ranges of up to 1 km and beyond between two similar devices without exceeding legal emission limits.[16][17][18]

While the Bluetooth Core Specification does mandate minimal for range, the range of the technology is application-specific and not limited. Manufacturers may tune their implementations to the range needed for individual use cases.

Bluetooth profiles

Main article: Bluetooth profile

To use Bluetooth wireless technology, a device has to be able to interpret certain Bluetooth profiles, which are definitions of possible applications and specify general behaviours that Bluetooth enabled devices use to communicate with other Bluetooth devices. These profiles include settings to parametrize and to control the communication from start. Adherence to profiles saves the time for transmitting the parameters anew before the bi-directional link becomes effective. There are a wide range of Bluetooth profiles that describe many different types of applications or use cases for devices.[19][20]

List of applications

A typical Bluetooth mobile phone headset.

- Wireless control of and communication between a mobile phone and a handsfree headset. This was one of the earliest applications to become popular.

- Wireless control of and communication between a mobile phone and a Bluetooth compatible car stereo system.
- Wireless control of and communication with tablets and speakers such as iPad and Android devices.
- Wireless Bluetooth headset and Intercom. Idiomatically, a headset is sometimes called "a Bluetooth".
- Wireless networking between PCs in a confined space and where little bandwidth is required.
- Wireless communication with PC input and output devices, the most common being the mouse, keyboard and printer.
- Transfer of files, contact details, calendar appointments, and reminders between devices with OBEX.
- Replacement of previous wired RS-232 serial communications in test equipment, GPS receivers, medical equipment, bar code scanners, and traffic control devices.
- For controls where infrared was often used.
- For low bandwidth applications where higher USB bandwidth is not required and cable-free connection desired.
- Sending small advertisements from Bluetooth-enabled advertising hoardings to other, discoverable, Bluetooth devices.[21]
- Wireless bridge between two Industrial Ethernet (e.g., PROFINET) networks.
- Three seventh and eighth generation game consoles, Nintendo's Wii.[22] and Sony's PlayStation 3, use Bluetooth for their respective wireless controllers.
- Dial-up internet access on personal computers or PDAs using a data-capable mobile phone as a wireless modem.
- Short range transmission of health sensor data from medical devices to mobile phone, set-top box or dedicated telehealth devices.[23]
- Allowing a DECT phone to ring and answer calls on behalf of a nearby mobile phone.
- Real-time location systems (RTLS), are used to track and identify the location of objects in real-time using “Nodes” or “tags” attached to, or embedded in the objects tracked, and

"Readers" that receive and process the wireless signals from these tags to determine their locations.[24]

- Personal security application on mobile phones for prevention of theft or loss of items. The protected item has a Bluetooth marker (e.g., a tag) that is in constant communication with the phone. If the connection is broken (the marker is out of range of the phone) then an alarm is raised. This can also be used as a man overboard alarm. A product using this technology has been available since 2009.[25]
- Calgary, Alberta, Canada's Roads Traffic division uses data collected from travelers' Bluetooth devices to predict travel times and road congestion for motorists.[26]